

ON OPTIMAL RETIREMENT

PHILIP A. ERNST,*
DEAN P. FOSTER* AND
LARRY A. SHEPP,* *University of Pennsylvania*

Abstract

We pose an optimal control problem arising in a perhaps new model for retirement investing. Given a control function f and our current net worth $X(t)$ for any t , we invest an amount $f(X(t))$ in the market. We need a fortune of M ‘superdollars’ to retire and want to retire as early as possible. We model our change in net worth over each infinitesimal time interval by the Itô process $dX(t) = (1 + f(X(t))) dt + f(X(t)) dW(t)$. We show how to choose the optimal $f = f_0$ and show that the choice of f_0 is optimal among all nonanticipative investment strategies, not just among Markovian ones.

Keywords: Retirement; optimal control problem; Itô process

2010 Mathematics Subject Classification: Primary 60H10
Secondary 60J60

1. Introduction

We begin by discussing the rationale and assumptions underlying the process chosen to model our change in wealth over each infinitesimal time interval. The model employs deflated (‘constant’) dollars and assumes that the investor borrows at the risk-free rate. The optimal control function ensures that the investor will only be able to borrow an amount of money such that, with probability 1, $X(t) \geq 0$ for all t . Like [2], the model assumes that the investor can either put money into or take out money out of the market in continuous time and that there is no transaction fee for doing so.

The investor chooses to invest solely in the ‘Sharpe asset’, explicitly defined in [3, p. 12]. In [3] it is proved that the Sharpe asset has close to unit variance (this is why $\sigma = 1$ in the volatility term $\sigma f(X(t)) dW(t)$). In formalizing the Itô process, we first normalize our unit of time so that the Sharpe asset is expected to return 100 percent over a unit time interval. This normalization forces $r = 1$ in the term $rf(X(t)) dt$. We proceed to normalize salary. Our monetary unit, which we will call a ‘superdollar’, is normalized such that we have a steady income of dt in each time interval of length dt . We need a fortune of M superdollars to retire and wish to retire as early as possible. See [8] and [9] for broader economic discussion of retirement processes.

The solution to this optimal control problem in retirement investing is especially interesting because it involves what seems to be a new phenomenon in boundary behavior, where a process hits a boundary and reflects ‘softly’ from it, without the need for local time as in reflecting Brownian motion. We show that the reflection takes place by using what seems to be a new approach to stochastic differential equations which avoids the clock-changing methods of [4] and [7]. Namely, in this approach we start by defining a particular ‘unit’ diffusion (i.e. where

Received 12 April 2013; revision received 10 June 2013.

* Postal address: The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104, USA.

the diffusion coefficient is identically unity) which is easier to construct using Picard’s method than a general diffusion would be. If we then take a particular monotonic function of the unit diffusion to construct the diffusion which is obtained as the answer f_0 (in Section 3), where the monotone function is nonnegative everywhere, it follows that the final diffusion is also always nonnegative.

We also consider a more general model of retirement investing where the diffusion coefficient, f , is replaced by Af^α . The general case reveals that the optimal investment strategy as well as the expected time until retirement are very strongly dependent in interesting ways on the particular model used.

2. Formal model and approach

Formally, the problem is stated as follows. Given $0 \leq x \leq M$ and a Brownian motion, $W(t)$, $t \geq 0$, find a nonanticipating process, $f(t)$, $t \geq 0$, so that if τ_M^f is the hitting time of M of the Itô process $X(t) = X^f(t)$, $t \geq 0$, with $X(0) = x$, and

$$dX(t) = (1 + f(X(t))) dt + f(X(t)) dW(t)$$

then $V(x; f) = \mathbb{E}_x[\tau_M^f]$ is a minimum over all such allowable f . Note that we are assuming that the state space for this optimal control problem is the right half-line so that we do not allow negative values of $X(t)$. Later, we will prove that, in our setup, $X(t)$ cannot be negative. If $X(t) < 0$ for some $t < \tau_M^f$, which is possible if $f(t)$ is bounded away from 0 and is also bounded, then there is a need to define what happens if the investor is in debt; we assume the game is over in this case and that $\tau_M^f = \infty$ so that with our definition we do not achieve a finite expectation much less a minimum. Other definitions allow for borrowing additional capital, but our model assumes that we are extremely adverse to being in debt.

We will show that, under this assumption, the optimum control, f_0 , exists and is unique. Any reasonable person would guess that $f(t) = f(X(t))$, i.e. that the optimal f is ‘Markovian’ in the sense that the optimal strategy depends only on the present fortune. But even if we guess that f should be Markovian, how do we learn which particular f is best? There is a nice way, involving a lot of nice guessing. Once one guesses f the proof that it is optimal is routine crank turning by martingale theory as we will see.

3. Formal statement of results

To get lower bounds on $V(x) = \inf_f V(x; f)$, one needs to find, in the usual way, a function, $\bar{V}(x)$, with $\bar{V}(M) = 0$, for which, for any f , the process, $Y(t) = t + \bar{V}(X^f(t))$ is a submartingale. If this is the case then we have from optional sampling, $\mathbb{E}_x[Y(\tau_M^f)] \geq Y(0)$. This gives, for any f and $0 \leq x \leq M$,

$$\mathbb{E}_x[\tau_M^f] = \mathbb{E}[Y(\tau_M^f)] \geq Y(0) = \bar{V}(x)$$

and, since this holds for any f and $0 \leq x \leq M$, we get $V(x) \geq \bar{V}(x)$.

Equality will hold for all x for the greatest lower bound, \bar{V} . The class of all such \bar{V} s is a convex class determined by the Itô inequalities defining a submartingale, which are that $\bar{V} \geq 0$ and that for all x and all f ,

$$\mathbb{E}[dY(t) \mid \mathcal{F}_t] = \bar{V}'(x)(1 + f) dt + \frac{f^2}{2} \bar{V}''(x) dt + dt \geq 0.$$

Since this must hold for all choices of f and all choices of $x = X(t)$ in $[0, \infty)$, and since this is quadratic in the real variable f (if this seems somewhat aggressive with respect to logic, recall

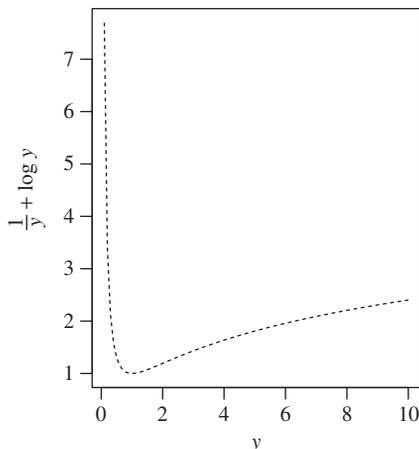


FIGURE 1: y against $(1/y) + \log y$.

that we are just using this reasoning for *guessing* the right \bar{V} . For any such \bar{V} we have for any f , $V(x; f) \geq \bar{V}(x)$, which gives us the lower bound, $\bar{V}(x)$ on $V(x)$. Which \bar{V} s satisfy the above submartingale condition?

Setting the derivative with respect to f equal to 0 we see that for each x we must have $\bar{V}''(x) > 0$, and then the minimum occurs at $f = f(x) = -\bar{V}'(x)/\bar{V}''(x)$. Putting this back into the submartingale inequality we need

$$1 + \bar{V}'(x) - \frac{1}{2} \frac{(\bar{V}')^2(x)}{\bar{V}''(x)} \geq 0.$$

For the best f , we need equality to hold everywhere in the string of inequalities above so that we would choose \bar{V} to satisfy the last inequality with equality throughout. If we set $g(x) = -\bar{V}'(x) \geq 0$ then we seek g to satisfy

$$g'(x) \left(\frac{1}{g(x)} - \frac{1}{g^2(x)} \right) \equiv \frac{1}{2}.$$

Integrating, we have, for some integration constant c ,

$$\frac{1}{g(x)} + \log(g(x)) = \frac{x + c}{2}.$$

Since the left-hand side is of the form $1/y + \log y \geq 1$ for all $y > 0$, it is tempting to choose $c = 2$ since this makes the right-hand side greater than or equal to 1 for $x \geq 0$ (it is our right to do this, since we are just guessing). We have almost arrived at a guess for the best \bar{V} , namely, we just have to solve the last equation for $g(x) = -\bar{V}'(x)$, and then \bar{V} is determined because we have $\bar{V}(M) = 0$.

A plot of y against $1/y + \log y$ is given in Figure 1 which shows that the inverse function defining $g(x)$ by

$$\frac{1}{g(x)} + \log(g(x)) = 1 + \frac{x}{2}$$

is not unique since the inverse is not one-to-one.

Which one do we use, the left-hand side branch or the right-hand side branch of Figure 1 to define $g(x)$ for each $x \in [0, M]$? Recall that we must have $g'(x) = -\bar{V}''(x) < 0$, so we guess

to use the left-hand side to determine $g(x)$. There is then clearly a unique solution $g(x)$ and we declare this as our guess at $g(x) = -\bar{V}'(x)$.

Using the condition $\bar{V}(M) = 0$, we have

$$\bar{V}(x) = - \int_x^M \bar{V}'(u) du = \int_x^M g(u) du.$$

We have already set up the proof that $\bar{V}(x) \equiv V(x)$. We have also seen that any optimal choice of $f = f_0$ must satisfy $f(x) = -\bar{V}'(x)/\bar{V}''(x)$, which we can express in terms of the $g(x)$ we have already defined. We have seen that V' can be expressed in terms of g , and so we obtain

$$f_0(x) = \frac{g(x)}{-g'(x)} = 2 \frac{1 - g(x)}{g(x)}.$$

Near $x = 0$, we can easily show through Maclaurin expansions for $1/(1 - \sqrt{x})$ and $\log(1 - \sqrt{x})$ that $g(x) \sim 1 - \sqrt{x}$, and so it follows that

$$f_0(x) = 2 \left(\frac{1}{g(x)} - 1 \right) \sim 2\sqrt{x}.$$

For an approximation of $g(x)$ when x is large, recall that we have defined $g(x)$ for $x \geq 0$ as the smaller of the two solutions of

$$\frac{1}{g(x)} + \log(g(x)) = 1 + \frac{x}{2}. \tag{3.1}$$

We next define

$$a = \frac{1}{g(x)} \quad \text{and} \quad b = 1 + \frac{x}{2}.$$

Rewriting (3.1) in terms of a and b , we obtain

$$a - \log a = b. \tag{3.2}$$

We now need an approximation for a in terms of b . We guess

$$a \approx b + \log b. \tag{3.3}$$

Substituting our approximation into (3.2) we obtain

$$b + \log b - \log(b + \log b) = b - \log\left(1 + \frac{\log b}{b}\right).$$

When b is large, $\log(b)/b$ is small, and, therefore, the left-hand side of (3.2) is very close to b . Substituting our definitions of a and b into (3.3), we obtain

$$\frac{1}{g(x)} \approx 1 + \frac{x}{2} + \log\left(1 + \frac{x}{2}\right),$$

or, equivalently,

$$g(x) \approx \frac{1}{1 + x/2 + \log(1 + x/2)}, \quad f_0(x) = 2 \left(\frac{1}{g(x)} - 1 \right) \approx x + 2 \log\left(1 + \frac{x}{2}\right).$$

To complete that proof that this $f = f_0$ is optimal, with this choice of $f = f_0$ and \bar{V} , the inequalities now hold for every other choice of f given that $Y^f(t) = t + \bar{V}^f(X^f(t))$ is a submartingale. It follows that $V(x) \geq \bar{V}(x)$ for all $x > 0$. We need to show that this equality

holds for the $f = f_0$ given above. It is true that in this case the submartingale is a *local* martingale, except possibly at 0. We need to show the equality $\mathbb{E}[Y(\tau_M)] = Y(0)$ holds where Y is the process $Y(t) = t + \bar{V}(X_0^f(t))$. It is enough to prove that $\tau_M < \infty$ with probability 1. The difficulty is that the process $X(t) = X^{f_0}(t)$ hits 0 uncountably many times with positive probability starting from any $0 \leq x < M$. How do we know that X cannot take negative values? When $X(t) = 0$ the unit drift moves it to the right, but how do we know that the term $f_0(X(t)) dW(t)$ does not cause the process to reach the negative half-line, or to get stuck at 0? Intuitively, each time the process hits 0, imagine that $f(x)$ is turned off so that $f(x) = 0$ for $0 \leq x \leq \varepsilon$. There is still a unit drift present so that the process takes time ε to reach the point $x = \varepsilon$. The probability starting at ε that the process hits M before it reaches 0 again is easily seen to be $1 - c\varepsilon$. It seems to follow from this that the expected time to reach M , starting from any x , is finite and the conclusion seems to follow. However, there is a mystery as to how the process pushes off from 0 and we will now resolve this point.

It seems remarkable that the process X^{f_0} behaves as if there is a reflecting barrier at 0. It does not pass through 0 to the negative half-axis because $f_0(0) = 0$. This means that it slows down as it gets near 0 but, unlike the Black–Scholes process [1], it actually hits 0. The drift of the process X^{f_0} is $1 + f(0) = 1$, so that it then moves away from 0 but it hits 0 uncountably many times (if it hits it once), just as the reflecting Brownian process $|W(t)|$ does, because the set of zeros of $X(t)$ is a perfect set.

It is instructive to consider a closely related reflecting process,

$$dX(t) = dt + 2\sqrt{X(t)} dW(t).$$

How does one prove that $X(t) \geq 0$ for all t ? This appears to be difficult because $X(t) = 0$ uncountably many times until one realizes that $X(t)$ is simply $W^2(t)$. Now it is indeed intuitive that $X(t)$ never goes negative, but we will now prove this rigorously.

Given any $A > 0$, consider the stochastic differential equation (SDE)

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad t > 0, X_0 = x_0 \in (0, A). \tag{3.4}$$

Theorem 3.1. *The process X is nonnegative.*

Proof. In our case

$$b(x) := 1 + 2\sqrt{x}, \quad \sigma(x) := 2\sqrt{x}, \quad \text{for } x \geq 0.$$

We assume that SDE (3.4) has a strong solution X , according to Definition 2.1 of [5, p. 285]. Its nonnegativity comes from Proposition 2.18 of [5, p. 293], where $X^{(1)} \equiv 0$ satisfies (almost surely (a.s.)) the SDE

$$X_t^{(1)} = X_0^{(1)} + \int_0^t b_1(X_s^{(1)}) ds + \int_0^t \sigma(X_s^{(1)}) dW_s, \quad t \geq 0,$$

for $X_0^{(1)} = 0$ and $b_1(\cdot) \equiv 0$, and $X^{(2)} \equiv X$ satisfies a.s. the SDE

$$X_t^{(2)} = X_0^{(2)} + \int_0^t b_2(X_s^{(2)}) ds + \int_0^t \sigma(X_s^{(2)}) dW_s, \quad t \geq 0,$$

for $X_0^{(2)} = X_0 = x_0$ and $b_2(\cdot) = b(\cdot)$. Assumptions (i) and (iii)–(v) of this proposition are easily checked (in the last one, $b_1(\cdot)$ satisfies condition (2.23) of [5]). Hence, it remains only

to show assumption (ii); i.e. for every $x, y \geq 0$ the diffusion coefficient of SDE (3.4) satisfies the condition

$$|\sigma(x) - \sigma(y)| \leq h(|x - y|) \tag{3.5}$$

for some strictly increasing function $h: [0, \infty) \rightarrow [0, \infty]$, such that $h(0) = 0$ and

$$\int_{(0, \varepsilon)} h^{-2}(u) \, du = \infty \quad \text{for all } \varepsilon > 0. \tag{3.6}$$

Indeed, for the strictly increasing function $h(x) = 2\sqrt{x}$ we have $h(0) = 0$. Condition (3.5) becomes

$$|2\sqrt{x} - 2\sqrt{y}| \leq 2\sqrt{|x - y|} \quad \text{if and only if } 0 \leq |x - y| \leq \sqrt{xy},$$

which holds for every $x, y \geq 0$, and

$$\int_{(0, \varepsilon)} \frac{du}{h^2(u)} = \frac{1}{4} \int_{(0, \varepsilon)} \frac{du}{u} = \frac{1}{4} [\ln u]_0^\varepsilon = \infty$$

for every $\varepsilon > 0$ that justifies condition (3.6). As a result, the above proposition gives

$$\mathbb{P}[X_t \geq 0 \text{ for all } t \geq 0] = \mathbb{P}[X_t^{(2)} \geq X_t^{(1)} \text{ for all } t \geq 0] = 1.$$

Theorem 3.2. *SDE (3.4) has a solution and it is unique.*

Proof. Thus far, we have proved that $X_t \geq 0$ for all $t \geq 0$, a.s. Thus, we now show that there exists a weak solution in the interval $I = (0, \infty)$ of SDE (3.4) up to the exit time

$$S := \inf\{t \geq 0; X_t \notin I\}, \tag{3.7}$$

according to Definition 5.20 of [5, p. 343]. This follows from Theorem 5.15 of [5, p. 341], where we need to verify that the coefficients of SDE (3.4) satisfy the assumptions of nondegeneracy,

$$\sigma^2(x) > 0 \quad \text{for all } x \in I,$$

and local integrability,

$$\text{for all } x \in I, \text{ there exists } \varepsilon > 0 \text{ such that } \int_{x-\varepsilon}^{x+\varepsilon} \frac{1 + |b(y)|}{\sigma^2(y)} \, dy < \infty.$$

Indeed, the former assumption holds trivially, while for the latter we may consider an $x > 0$ and take $\varepsilon := x/2 > 0$ to obtain

$$\int_{x-\varepsilon}^{x+\varepsilon} \frac{1 + |b(y)|}{\sigma^2(y)} \, dy = \int_{x-x/2}^{x+x/2} \frac{2 + 2\sqrt{y}}{4y} \, dy = \left[\frac{1}{2} \ln y + \sqrt{y} \right]_{y=x/2}^{y=3x/2} < \infty.$$

We show next that strong uniqueness in the interval $I = (0, \infty)$ holds for SDE (3.4) up to the exit time S , according to Definition 2.3 of [5, p. 286]. This is a consequence of Theorem 2.5 of [5, p. 287], after we check that the coefficients of SDE (3.4) are locally Lipschitz on I ; in other words, for every closed set $J \subseteq I$, there exists a constant $K_J > 0$ such that

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq K_J|x - y| \quad \text{for all } x, y \in J.$$

Indeed, let $J = [a, b]$ with $0 < a < b$. Then, for every $a \leq x, y \leq b$, we have

$$|b(x) - b(y)| = |\sigma(x) - \sigma(y)| = 2|\sqrt{x} - \sqrt{y}| = 2\frac{|x - y|}{\sqrt{x} + \sqrt{y}} \leq 2\frac{|x - y|}{\sqrt{a}} \leq 2\frac{|x - y|}{\sqrt{a}}.$$

Finally, owing to Corollary 3.23 of [5, p. 310], weak existence and strong uniqueness imply strong existence in the interval $I = (0, \infty)$ for a solution of SDE (3.4) up to the exit time S , according to Definition 2.1 of [5, p. 285].

4. Boundary behavior of the process X on $(0, A)$

Let X be the unique strong solution of the SDE (3.4) up to the exit time S of (3.7). Note that the assumption $X_0 = x_0 \in (0, A)$ guarantees that $S > 0$ a.s. Since the assumptions of nondegeneracy and local integrability above are satisfied by the coefficients of SDE (3.4), we compute the scale function

$$S(x) := \int_1^x s(\xi) \, d\xi, \quad x \in I,$$

where

$$s(\xi) := \exp\left(-2 \int_1^\xi \frac{b(\zeta)}{\sigma^2(\zeta)} \, d\zeta\right) = \exp\left(-2 \int_1^\xi \frac{(1 + 2\sqrt{\zeta})}{(2\sqrt{\zeta})^2} \, d\zeta\right) = \frac{1}{\sqrt{\xi}} e^{2(1-\sqrt{\xi})}$$

for all $\xi \in I$. Thus,

$$S(x) = \int_1^x s(\xi) \, d\xi = \int_1^x \frac{1}{\sqrt{\xi}} e^{2(1-\sqrt{\xi})} \, d\xi = 1 - e^{2(1-\sqrt{x})} \quad \text{for all } x \in I,$$

which is the same with the function $p(\cdot)$ given in Equation (5.42) of [5, p. 339]. We may also compute the speed density

$$m(x) := \frac{2}{\sigma^2(x)s(x)} = \frac{2}{4x(1/\sqrt{x})e^{2(1-\sqrt{x})}} = \frac{e^{2(\sqrt{x}-1)}}{2\sqrt{x}} \quad \text{for all } x \in I,$$

and accordingly the speed measure

$$\begin{aligned} M[a, b] &:= \frac{1}{2} \int_a^b m(x) \, dx \\ &= \int_a^b \frac{e^{2(\sqrt{x}-1)}}{4\sqrt{x}} \, dx \\ &= \frac{1}{4} \int_{2(\sqrt{a}-1)}^{2(\sqrt{b}-1)} e^u \, du \\ &= \frac{1}{4} (e^{2(\sqrt{b}-1)} - e^{2(\sqrt{a}-1)}) \quad \text{for all closed sets } [a, b] \subseteq I. \end{aligned}$$

Recalling that $X_0 = x_0 \in (0, A) \subseteq I$, we define the exit times

$$T_{a,A} := \inf\{t \geq 0; X_t \notin (a, A)\} \quad \text{for } 0 < a < A, \quad \tau := \inf\{t \geq 0; X_t \notin (0, A)\}$$

of the process X from the intervals (a, A) and $(0, A)$, respectively, where $\mathbb{P}[T_{a,A} > 0] = \mathbb{P}[\tau > 0] = 1$, and the relationships given in Equation (5.61) of [5, p. 344] yield

$$\mathbb{P}[X_{T_{a,A}} = a] = \frac{S(A) - S(x_0)}{S(A) - S(a)} \quad \text{and} \quad \mathbb{P}[X_{T_{a,A}} = A] = \frac{S(x_0) - S(a)}{S(A) - S(a)}.$$

In order to obtain information about the behavior of X near the endpoints of the interval $(0, A) \subseteq I$, we easily see that $S(0^+) = 1 - e^2 > -\infty$.

We may follow the reasoning of Proposition 5.22 of [5, p. 345] to take limits in the above expressions and eventually get the exit probabilities from the interval $(0, A)$ as

$$\begin{aligned} \mathbb{P}\left[\lim_{t \uparrow \tau} = 0\right] &= 1 - \mathbb{P}\left[\lim_{t \uparrow \tau} = A\right] \\ &= \frac{S(A) - S(x_0)}{S(A) - S(0+)} \\ &= \frac{-e^{2(1-\sqrt{A})} + e^{2(1-\sqrt{x_0})}}{-e^{2(1-\sqrt{A})} + e^2} \\ &= \frac{e^{2(\sqrt{A}-\sqrt{x_0})} - 1}{e^{2\sqrt{A}} - 1}. \end{aligned}$$

We argue next that exit from the interval $(0, A)$ occurs a.s. in a finite expected time, i.e. $\mathbb{P}[\tau < \infty] = 1$ and $\mathbb{E}[\tau] < \infty$.

To this end, it is sufficient to proceed as in the proof of Proposition 5.32(i) of [5, p. 350] with respect to the function given in Equation (5.65) of [5, p. 347]:

$$\begin{aligned} v(x) &= \int_1^x s(y) \int_1^y m(z) dz dy \\ &= \int_1^x (S(x) - S(y))m(y) dy \\ &= S(x) \int_1^x m(y) dy - \int_1^x S(y)m(y) dy \\ &= [1 - e^{2(1-\sqrt{x})}] \int_1^x \frac{e^{2(\sqrt{y}-1)}}{2\sqrt{y}} dy - \int_1^x [1 - e^{2(1-\sqrt{y})}] \frac{e^{2(\sqrt{y}-1)}}{2\sqrt{y}} dy \\ &= -e^{2(1-\sqrt{x})} \int_1^x \frac{e^{2(\sqrt{y}-1)}}{2\sqrt{y}} dy + \int_1^x \frac{1}{2\sqrt{y}} dy \\ &= -\frac{1}{2}(1 - e^{2(1-\sqrt{x})}) + \sqrt{x} - 1, \quad x \in I. \end{aligned}$$

Finally, observe that $v(0^+) = 1 - \frac{1}{2}(1 - e^2) - 1 < \infty$ and $v(A) = -\frac{1}{2}(1 - e^{2(1-\sqrt{A})}) + \sqrt{A} - 1 < \infty$.

We conclude the argument according to Remark 5.33 of [5, p. 350]. In particular, we also have finiteness of $S(0+)$ and $S(A)$ (thanks to Problem 5.27 of [5, p. 348]), and we can define, by analogy with Equation (5.52) of [5, p. 343], the Green function

$$G_{0,A}(x, y) := \frac{[S(x \wedge y) - S(0+)] [S(A) - S(x \vee y)]}{S(A) - S(0+)} \quad \text{for all } (x, y) \in (0, A]^2.$$

Then the counterpart of Equation (5.22) of [5, p.343] is defined by

$$\begin{aligned} M_{0,A}(x) &:= \int_0^A G_{0,A}(x, y)m(y) dy \\ &= - \int_0^x [S(x) - S(y)]m(y) dy + \frac{S(x) - S(0+)}{S(A) - S(0+)} \int_0^A [S(A) - S(y)]m(y) dy \end{aligned}$$

for all $x \in (0, A]$, and satisfies $M_{0,A}(0+) = M_{0,A}(A) = 0$ as well as Equation (5.53) of [5, p. 343]. The same procedure that led to Equation (5.59) of [5, p. 343] now gives $\mathbb{E}[\tau] = M(x_0) < \infty$. Consequently, the unique strong solution X of the SDE (3.4) exits from the interval $(0, A)$ at each endpoint with positive probability and in finite expected time.

Theorem 4.1. *The process X is reflecting at 0.*

Proof. We want to ensure that the unique strong solution X of the SDE (3.4) will eventually exit a.s. from the interval $(0, A)$ at A . To this end, we will specify the behavior of the process X at the exit point 0. According to the ‘regular boundary’ characterization of [6, pp. 232–233], a diffusion process can both enter and leave from a regular boundary. Since

$$\begin{aligned} S(0, x_0) &:= S(x_0) - S(0+) = 1 - e^{2(1-\sqrt{x_0})} - [1 - e^{2(1-\sqrt{0})}] = e^2(1 - e^{-2\sqrt{x_0}}) < \infty, \\ M(0, x_0) &:= M[0+, x_0] \\ &= M(x_0) - M(0+) \\ &= \frac{1}{4}(e^{2(\sqrt{x_0}-1)} - e^{2(\sqrt{0}-1)}) \\ &= \frac{1}{4e^2}(e^{2\sqrt{x_0}} - 1) \\ &< \infty, \end{aligned}$$

there is a regular boundary at 0 and we may assign a speed $M[\{0\}]$ to itself. Indeed,

$$M(0, x] = \lim_{a \downarrow 0} \int_a^x m(y) dy = \lim_{a \downarrow 0} \frac{1}{4}(e^{2(\sqrt{x}-1)} - e^{2(\sqrt{a}-1)}) = \frac{1}{4e^2}[e^{2\sqrt{x}} - 1] \tag{4.1}$$

and

$$M[\{0\}] = \lim_{x \downarrow 0} M(0, x] = \lim_{x \downarrow 0} \frac{1}{4e^2}(e^{2\sqrt{x}} - 1) = 0,$$

which implies that 0 is a reflecting boundary point. Therefore, if the process X exits firstly from the interval $(0, A)$ at the regular boundary 0 then it is reflected back to its interior and keeps doing this until it eventually exits at the regular boundary A .

Theorem 4.2. *X reflects off 0 even though there is no local time in its Itô representation.*

Proof. It is interesting that X reflects off 0 even though there is no local time in its Itô representation. The way we prove that this makes sense seems to be new and it seems to shed much light on the basic existence proofs of Itô theory. In Itô theory, Picard iteration easily shows that a diffusion $X = X(t, \omega)$ with a unit diffusion coefficient,

$$dX(t) = A(X(t)) dt + dW(t), \quad t \geq 0, X(0) = x_0,$$

can be constructed on any space on which a process $W(t, \omega)$ with continuous sample paths is available, path by path, if A has bounded difference quotients. It is then possible to construct a very general diffusion $Y(t)$, satisfying $dY(t) = a(Y(t)) dt + b(Y(t)) dW(t)$ by constructing the unit diffusion $X(t)$ with appropriate drift $A(x)$ and then setting $Y(t) = g(X(t))$ for an appropriate function $g = g(x)$. It is easy to check that to make Y into an $a(y), b(y)$ diffusion, we must choose $A(x)$ and $g(x)$ as follows:

$$g'(x) = b(g(x)) \quad \text{and} \quad A(x) = \frac{a(g(x)) - b'(g(x))g'(x)/2}{b(g(x))}.$$

If we do this for the case of f_0 which solves the problem, we see that $g(x) \geq 0$, so it follows that the diffusion never goes negative and reflects softly at $x = 0$, just as in the case of $W^2(t), t \geq 0$.

It seems remarkable that g is defined only by the diffusion coefficient, b . But this is somewhat illusory because A enters as well in that the unit diffusion process, X , depends on A , and if A does not satisfy a Lipschitz condition then the domain of the X diffusion may be a subset of

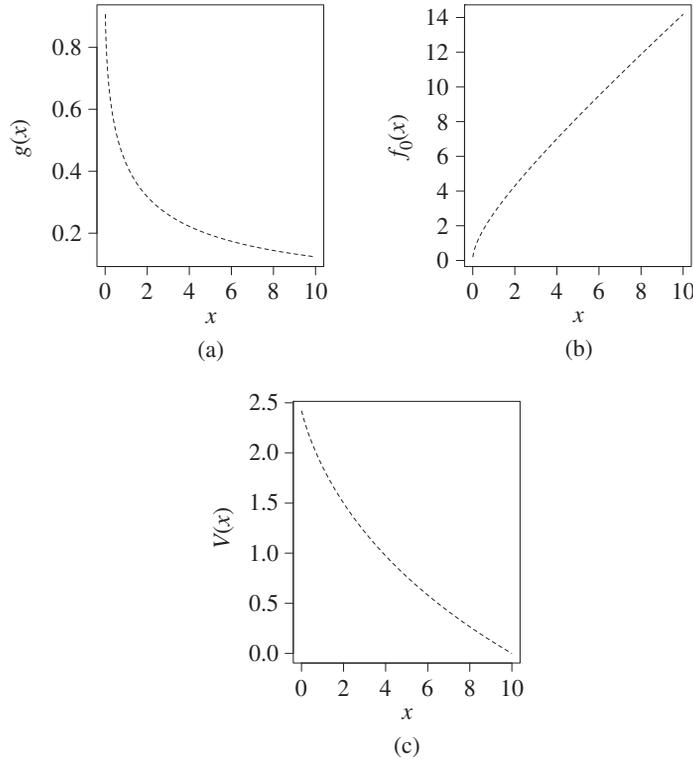


FIGURE 2: Plots of (a) $g = g(x)$ for $1/g + \log(g) = 1 + x/2$, (b) the optimal $f = f_0$, $f(x) = g(x)/g'(x)$, and (c) the optimal payoff, $V_M^f(x)$, $M = 10$, $f = f_0$.

the whole line. Let us look at the case

$$dY(t) = dt + c\sqrt{|Y(t)|} dW(t), \quad t \geq 0, Y(0) = 0,$$

in more detail since this is an example which is very similar to the optimal control process of the paper. We can carry out the steps of the determination of B , g , and A , above to see that

$$B(y) = \int_0^y \frac{du}{c\sqrt{u}} = \frac{2}{c}\sqrt{|y|\text{sgn}(y)}, \quad g(x) = \frac{c^2x^2}{4}\text{sgn}(x), \quad A(x) = \frac{2(1 - (c^2/4)\text{sgn}(x))}{c^2|x|}.$$

We see there is an infinite singularity in A at $x = 0$ except if $c = 2$. If $0 < c < 2$ the process X never hits 0, but if $c > 2$ then $X(t)$ cannot be defined after it hits 0, at least not by the diffusion equation above.

This completes the proof that, for the optimal investment strategy, $f = f_0$, our fortune reaches M in a finite time with minimum expected value. It is remarkable that we go broke repeatedly with positive probability before achieving our goal.

Remark 4.1. It is often remarked of some rich people that because they were ‘aggressive, they went into bankruptcy several times before making it’. Somehow, mathematics seems to have already been aware of this common observation! Note that it is always true that $V(x) \leq M - x$ since an investor can always choose $f \equiv 0$ and ‘save the way to retirement’.

A graph of $g(x) = -V'(x)$ is given in Figure 2(a), a graph of the optimal investment strategy, f_0 , is given in Figure 2(b), and a graph of the optimal payoff, $V(x)$, is given in Figure 2(c).

5. Generalization of the problem

A more general model for retirement than the one of Section 3, namely,

$$dX(t) = (1 + f(X(t))) dt + f(X(t)) dW(t), \quad X(0) = x,$$

would allow the diffusion term to be any fixed function of $f(X(t))$ rather than simply $f(X(t))$ itself. The choice made above was also used in the Black–Scholes–Samuelson model for stock prices which was based on the argument that doubling an investment empirically seems to double the volatility, but this is a crude argument and other possibilities seem to be worth exploring. We propose considering the more general model

$$dX(t) = (1 + f(X(t)) dt + \phi(f(X(t))) dW(t), \quad X(0) = x,$$

where $\phi(u)$ is any increasing function. For tractability, we will restrict the discussion to the particular forms $\phi(u) = Au^\alpha$, where $A > 0$ and $\alpha > 0$ are parameters.

The same method of proof shows that, for $\alpha = 1$, as before, but using general A , we have

$$g(x; A; \alpha = 1) = -V'(x; A; \alpha = 1) = g\left(\frac{x}{A^2}; A = 1; \alpha = 1\right).$$

It follows that

$$V(x; A; \alpha = 1) = A^2 V\left(\frac{x}{A^2}; A = 1; M = \frac{M}{A^2}\right).$$

This is as expected; the original model used f as both the drift and the diffusion parameter because one could scale time to make the diffusion equal to 1 in appropriate time units. However, if one wants to *compare* models then the parameter A must be retained. If one does this, one sees that

$$g(x; A; \alpha = 1) \rightarrow g(0; A = 1; \alpha = 1) \equiv 1,$$

so that

$$\lim_{A \rightarrow \infty} V(x; A; \alpha = 1) = \int_x^M du = M - x.$$

The conclusion is that if the investor has the choice of investing in a risky market or instead to be conservative by saving his/her salary without investing, then the conservative strategy is asymptotically (as $A \rightarrow \infty$) superior even though the resulting time to retirement is $M - x$, which is the maximum delay among all models since there is no advantage to investment. The conservative investment advice to avoid risk is usually given to older investors; our modeling assumptions conclusions bear this out, even for young investors in the limit as risk gets very large.

We next consider the case $\frac{1}{2} \leq \alpha < 1$. Since the diffusion speed is smaller for $\alpha < 1$ than for $\alpha = 1$, one would think that as α decreases the expected time would increase and investing would be disadvantageous, but this is surprisingly not the case, as we see below. Moreover, we will show that, for $A = 1$ and $0 < \alpha < \frac{1}{2}$, one can find investment strategies that allow retirement in time ε , arbitrarily small. Another surprise is that there is a sharp discontinuity in $V(x, \alpha)$ as $\alpha \uparrow \frac{1}{2}$. We show that $V(x, \alpha) = 0$ for $\alpha < \frac{1}{2}$, but, as we see below,

$$V\left(x; \alpha = \frac{1}{2}\right) = \frac{1}{2}(e^{-2x} - e^{-2M}).$$

We turn to the $A > 0, \alpha \in [0.5, 1)$ problem: again, we define $V_M^f(x) = \mathbb{E}_x^f[\tau_M]$, and we want to find $V_M(x) = \inf_f V_M^f(x)$.

Again, we seek any function $\bar{V}(x)$ for which $Y(t) = \bar{V}(X^f(t)) + t$ is a submartingale for any choice of f . Any such $\bar{V}(x)$ will be a lower bound on $V(x)$ since we will again have

$$\mathbb{E}_x[\bar{V}(X^f(\tau_M))] + \mathbb{E}[\tau_M^f] \geq \mathbb{E}_x[\bar{V}(X^f(0))] = \bar{V}(x).$$

To be a submartingale, we need, in the same way as in the case $\alpha = 1$,

$$\bar{V}'(x)(1 + f) + \frac{A^2}{2}\bar{V}''(x)f^{2\alpha} + 1 \geq 0, \quad 0 \leq x \leq M, f \geq 0.$$

For fixed x , this is a minimum in f at the point

$$f(x) = f_0(x) = \left(\frac{-\bar{V}'(x)}{\bar{V}''(x)} \frac{1}{A^2\alpha} \right)^{1/(2\alpha-1)}.$$

Requiring that Y be a martingale for the best choice of $f = f_0$ above gives an ordinary differential equation (ODE) for $g(x) = g(x, A, \alpha) = -\bar{V}'(x, A, \alpha)$. After a calculation, very similar to the one above for $A = \alpha = 1$, the ODE is

$$\frac{-g'(x)(1 - g(x))^{2\alpha-1}}{g^{2\alpha}(x)} = \frac{(1 - 1/2\alpha)^{2\alpha-1}}{A^2\alpha}.$$

Integrating gives

$$\int_{g(x)}^1 \frac{(1 - u)^{2\alpha-1}}{u^{2\alpha}} du = x \frac{(1 - 1/2\alpha)^{2\alpha-1}}{A^2\alpha} + c.$$

Again, we guess that $c = 0$ and we can solve for $g(x) \in [0, 1]$ for any $x \geq 0$. We see that, so long as $\alpha > 0.5$, there is no trouble. We can write (since $V(M) = 0$)

$$\begin{aligned} V(x) &= \int_x^M -V'(u) du \\ &= \int_x^M g(u) du \\ &= \int_x^M \frac{g(u)}{g'(u)} g'(u) du \\ &= (b(\alpha))^{-1} \int_{g(M)}^{g(x)} \left(\frac{1 - u}{u} \right)^{2\alpha-1} du, \end{aligned}$$

where

$$b(\alpha) = \frac{(1 - 1/2\alpha)^{2\alpha-1}}{\alpha A^2}.$$

Finally, $f_0 = f_0(x, A, \alpha)$ is given by

$$f_0(x) = \left(\frac{1}{g(x)} - 1 \right) \frac{1}{1 - 1/2\alpha}.$$

6. General model in the case of $\alpha < \frac{1}{2}$

The case $\alpha < \frac{1}{2}$ is especially interesting. The argument given for guessing the optimal \hat{V} breaks down.

Theorem 6.1. *There is no \hat{V} that will make $\hat{V}(X^f(t)) + t$ a submartingale for all choices of $f = f(x)$ except the trivial case $\hat{V}(x) \equiv 0$. This is the best lower bound that submartingale theory can provide, which leads one to suspect that $V(x; \alpha) \equiv 0$ for $\alpha < \frac{1}{2}$.*

Proof. Here we assume that $A = 1$. We need to find an f that makes the expected time to reach M arbitrarily small. Consider the investment strategy

$$f(x) = 0, \quad 0 \leq x \leq \varepsilon, \quad f(x) = c, \quad \varepsilon < x < M.$$

If we can find a function, $g = g(x) = g_{\varepsilon,c}(x)$, $0 \leq x \leq M$, for which $Y(t) = g(X_t) + t$ is a martingale and $g(M) = 0$, then the martingale theorem gives $g(x) = Y(0) = \mathbb{E}_x[Y(\tau_M)] = \mathbb{E}_x[\tau_M]$. Itô calculus states that g must be, for appropriate constants A' , B , and D ,

$$g(x) = \begin{cases} D - x, & 0 \leq x \leq \varepsilon, \\ \frac{M - x}{1 + c} + A' \left(\exp\left(-\frac{2(1 + c)}{c^{2\alpha}}x\right) - \exp\left(-\frac{2(1 + c)}{c^{2\alpha}}M\right) \right), & \varepsilon \leq x \leq M. \end{cases}$$

We determine from continuity at $x = \varepsilon$ of g and g' that

$$A' = \frac{c^{2\alpha+1}}{2(1 + c)^2} \exp\left(\frac{2(1 + c)}{c^{2\alpha}}\varepsilon\right),$$

$$D = \frac{M - \varepsilon}{1 + c} + A' \left(\exp\left(-\frac{2(1 + c)}{c^{2\alpha}}\varepsilon\right) - \exp\left(-\frac{2(1 + c)}{c^{2\alpha}}x\right) \right), \quad \varepsilon \leq x \leq M.$$

Finally, set $\varepsilon = 1/(1 + c)$, let $c \rightarrow \infty$, and verify that A' and D tend to 0 as $c \rightarrow \infty$ to show that $V(x; \alpha) \equiv 0$ for $\alpha < \frac{1}{2}$.

Remark 6.1. Under this model, one can retire in an arbitrarily small expected time. This should probably be interpreted as meaning that the model with $\alpha < \frac{1}{2}$ does not represent the real world; models should be looked at carefully and rejected if they do not conform to observation.

Acknowledgements

We thank Professor Lawrence Brown and Professor J. Michael Harrison for their invaluable input to many of the ideas in this paper.

References

- [1] BLACK, F. AND SCHOLES, M. (1973). The pricing of options and corporate liabilities. *J. Political Econom.* **81**, 637–654.
- [2] COVER, T. M. (1991). Universal portfolios. *Math. Finance* **1**, 1–29.
- [3] FOSTER, D. P., KAKADE, S. AND RONEN, O. (2007). Early retirement using leveraged investments. Preprint.
- [4] ITÔ, K. AND MCKEAN, H. P. (1965). *Diffusion Processes and Their Sample Paths*. Academic Press, New York.
- [5] KARATZAS, I. AND SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer, New York.
- [6] KARLIN, S. AND TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York.
- [7] MCKEAN, H. P. (1969). *Stochastic Integrals*. Academic Press, New York.
- [8] MERTON, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Rev. Econom. Statist.* **51**, 247–257.
- [9] MERTON, R. C. (1992). *Continuous-Time Finance*. Blackwell, MA.