

Optimal Purchasing of Deferred Income Annuities When Payout Yields are Mean-Reverting*

Huaxiong Huang¹, Moshe A. Milevsky², and Virginia R. Young³

¹York University, ²Schulich School of Business, York University and ³University of Michigan

Abstract

We determine the optimal lifecycle purchasing strategy for deferred income annuities (DIAs)—which are distinct from single-premium income annuities (SPIAs)—for an individual who wishes to maximize the expected utility of his/her annuity income at a fixed time in the future. In contrast to the vast portfolio-choice literature for SPIAs, we focus on the stochasticity of the DIA's payout yield and address concerns that rates are currently “too low” to justify irreversible annuitization. We assume a mean-reverting model for payout yields and show that a risk-neutral consumer who wishes to maximize his/her expected retirement income should wait until yields reach a threshold—which lies above historical averages—and then purchase the DIA in one lump sum. In contrast, a risk-averse consumer who is concerned the payout yield will remain below average for an extended period and worries about losing mortality credits while waiting, should employ a barrier purchasing strategy, as in the portfolio choice problem under transaction costs. We illustrate how this insight is applied in the context of annuitization. In fact, the optimal behavior of a risk-averse consumer resembles an asymmetric dollar-cost averaging strategy, with a portion of the DIA-budget spent even while payout rates are below historical averages. As part of our analysis we offer an easy-to-use asymptotic approximation for the optimal purchasing strategy (threshold) and provide some numerical examples to illustrate the concept.

JEL classification: G11, D9, C6, G22

1. Introduction

Similar to a Defined Benefit pension, a deferred income annuity (DIA) is no different from the centuries-old life-contingent annuity in which the insured pays a lump sum in exchange

* The first and second author acknowledge funding from the IFID Centre and thank CANNEX for access to its annuity quotation system. The first and second authors also have a financial relationship with CANNEX, which provided some preliminary DIA quotes. Research of the first author has been supported by NSERC; the second author by SSHRC; and third author by the Cecil J. and Ethel M. Nesbitt Professorship of Actuarial Mathematics.

for a guarantee of lifetime income. The pricing of life annuities has been well understood and established for over three centuries, ever since the classic work by Edmond Halley in his famous paper published in *Philosophical Transactions* in 1693.

In contrast to a single-premium income annuity (SPIA), a DIA—which is the focus in this article—is purchased at a relatively young age, and the income starting date is delayed until a relatively old age. The start date of a DIA is often decades away and well after the normal age of retirement; hence, the word deferred in its name. The growing appeal and popularity of DIAs is that they can serve as a pension substitute or replacement for consumers and retirees who do not have longevity insurance in the form of Defined Benefit pensions provided by employers.

A recent article in the *Wall Street Journal* on January 17 2015, touting the benefits of this form of longevity insurance, reported that total DIA sales in the USA during the year 2014 were \$3 billion USD. This volume of sales in 2014 was 35% higher than sales in 2013, which is astonishing growth for a staid insurance product.¹ DIA sales have been growing much faster than other forms of life insurance—certainly compared to SPIAs—hence, there is much interest in DIAs, separate from ordinary, immediate annuities.

Part of the reason for the growth of interest in DIAs is a recent decision by the US Treasury department (July 2014) to allow the purchase of DIAs inside tax-sheltered retirement plans and to exempt them from certain distributional requirements at age 70, as long as the annuity commences the income phase prior to age 85. This is also known as an Advanced Life Delayed Annuity (ALDA). A second announcement by the US Treasury (November 2014) to allow them as defaults within target-date and life-cycle funds (with certain provisos) has added to the interest among practitioners and even the public. A casual reader of the business press encounters articles about DIAs on a weekly basis.

Even though the pricing methodology underlying (new, popular) DIAs is the same as for (old, less popular) SPIAs, the fact that income payments are delayed for 10, 20, and possibly 30 years implies that pricing becomes extremely sensitive to the far-end of the term structure of interest rates. The duration of cash flows in a DIA is obviously much higher relative to a SPIA. And, whereas in early 2015 global interest rates are at historical lows, industry participants express a valid concern that this particular environment might not be the best time to purchase any long-term fixed-income instrument, including a DIA. Because a DIA is effectively a fixed-income product, its price will invariably decline—and its payout yield will increase—when interest rates revert to normal (historical) levels. A \$100 deposit today might guarantee a lifetime income of \$40 per year beginning in 20 years, but perhaps next year, or even next month, the same \$100 will generate \$50 of deferred lifetime income. So, why not wait?

This question “why not wait?” is exactly the point on which we focus in this article. Namely, we investigate the optimal strategy for purchasing DIAs in a stochastic pricing environment in which payout yields are mean reverting, driven by the assumption that interest rates are expected to slowly return to higher levels. We believe that neither the financial nor the actuarial insurance literature has dealt with this specific problem, and in this article we provide a detailed analysis—with some asymptotic approximations and numerical examples—for the optimal purchase policy.

1 These sales figures are according to Life Insurance Marketing Research Association (LIMRA), a company that tracks sales figures and trends in the insurance industry.

Here are some real-world numbers to put this problem in context. On December 24 2014, there were approximately a dozen insurance companies in the USA offering both DIAs and SPIAs for sale and quoting prices in the CANNEX online system (www.cannex.com). The top five quoting insurance companies offered to pay a 68-year-old male an average of \$600 per month, immediately, in exchange for a lump-sum premium of \$100,000. This insurance product is a SPIA, whose price is often expressed as an annualized payout yield of 7.25%. On the same date, 30-year US government bond yields were approximately 2.83%, according to the US Federal Reserve. The 450-basis-point difference between the two yields is due to (i) mortality credits and irreversibility of the purchase of a SPIA, (ii) credit default risk of the insurer leading to a higher discount rate for a SPIA, and (iii) different patterns of cash flows over time. Recall, the 30-year bond returns the original principal at maturity, whereas the principal in a life annuity is paid out and amortized over the remaining lifetime of the annuitant.

In contrast to a SPIA, the same \$100,000 premium at age 68 (for a male), would have guaranteed a payout yield of 17.77% (the average quoted by the top five insurance companies) if the income were deferred for 10 years to age 78. In other words, the \$100,000 would generate \$1,480 per month. In this article we use the symbol $\pi_t(x, y)$ to denote the payout yield at time t , for an individual who was age x at time 0, with income starting age of y . So, using this notation, $\pi_0(68, 78) = 0.1777$, in which time $t = 0$ is December 24 2014.

Finally, if the income were delayed by 20 years to age 88, the payout yield quoted was $\pi_0(68, 88) = 64.60\%$, which is \$5,380 per month for life. The reason the time-0 payout yield of a DIA is greater than the payout yield of a SPIA is that the individual is not receiving any income between ages x and y . Furthermore, the DIA pays the income over time, as does a SPIA, instead of in a lump sum as with a zero-coupon bond. Another way for bond mavens to “think” about a DIA is as a laddered portfolio of zero-coupon bonds, the first of which starts paying at age y and the last payment is made at the age of 120. But, if the individual dies, the remaining portfolio of bonds is lost, with a recovery rate of zero.

We call the entire surface $\pi_t(x, y)$, which traces out the payout yield, from immediate income ($y = x$) to delayed income ($y \gg x$), the term structure of DIA payout yields at time t . Figure 1 displays the term structure of annuity payout yields (APYs) at the most recent date for which data were available, on December 24 2014. It traces out the π -curves for a 68-year-old male and female, with income starting at various future dates ranging from $T = 0$ to $T = 20$ (ages $y = 68$ to $y = 88$). Once again, we selected this combination of ages because it is closest in spirit to the ages alluded to in the US Treasury announcement and would also involve the highest possible mortality credits. Our methodology, of course, can be applied to any age combination.

We remark that this surface changes from day to day and prices move as frequently as any bond traded in the market. For example, on December 24 2014, the highest payout yield quoted by the dozen companies offering DIAs was $\pi(68, 88) = 73.18\%$, but one month earlier, on November 24 2014, the highest payout yield—albeit not from the same company—was $\pi(68, 88) = 73.38\%$. This was a change of 20 basis points during the course of one month. During the same one-month period, the 30-year Treasury bond yield dropped from 3.01% to 2.83%, which is slightly less than 20 basis points. Also, we estimate the annualized volatility of $\pi_t(68, 88)$ for individual companies to be 3–12% per year, depending on how frequently a company updates its prices and how competitive it is in the

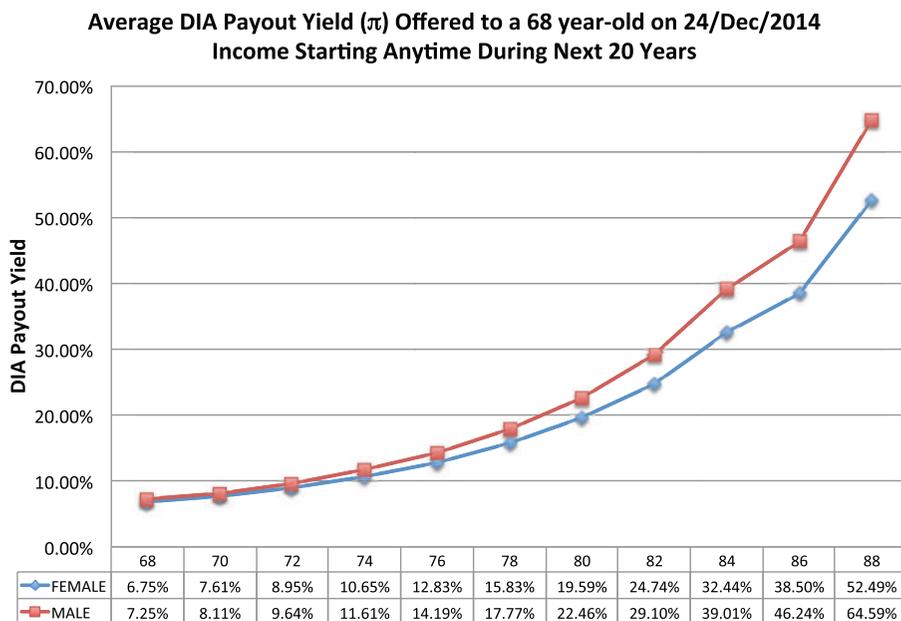


Figure 1. Average—across all quoting companies—of APYs $\pi_t(68, 68 + s)$ for $s = 0.20$, using the notation introduced in the article, where $t = 0$ is December 24 2014. Note that individual companies might quote rates that are 1–3% higher or lower than the average. Rates change daily.

marketplace. Although given the limited amount of data available on these (newer, delayed income) products, our estimates should be viewed as rather casual.

In fact, there are many other frictional factors that drive annuity quotes from day to day and we refer to the recent empirical work by [Kojien and Yogo \(2015\)](#) as well as [Charupat, Kamstra, and Milevsky \(2015\)](#) for more information about the microstructure of annuity prices.

This background sets the stage for the formal question we address in this article. Assume an individual has \$100,000 (or any other fixed sum) that he/she plans to allocate toward the purchase of a DIA during the next T years. What is the optimal dynamic policy he/she should employ for this purpose? After all, if-and-when interest rates increase over the next few years, DIA payout yields will improve as well. On the other hand, if not spent, the \$100,000 will miss out on the mortality credits embedded within the DIA's price while the consumer is waiting to annuitize. Moreover, although pricing (and interest) rates might revert to the mean over time, it is unclear how long it will take for these rates to return to historic norms. Different market participants have their own views on the half-life of the current low rates, so the speed of mean reversion is a factor as well.

We want to emphasize that our article differs from the prior annuity-timing and allocation literature, such as [Bayraktar and Young \(2009\)](#), [Milevsky and Young \(2007\)](#), [Stabile \(2006\)](#), [Kingston and Thorpe \(2005\)](#) or going back to the very early work by [Richard \(1975\)](#) and [Yaari \(1965\)](#), by abstracting from portfolio considerations. In other words, we do not address the optimal allocation to DIAs and SPIAs versus stocks and bonds, or whether a 42-year-old (e.g.) should have 20% or 40% or 0% allocated to DIAs. These are important issues that form part of a long-standing and growing literature, but they distract

from our main point which is mean-reverting payout yields.² We also do not address the much larger literature on the annuity puzzle and the growing behavioral research around the psychological barriers to annuitization.³

This article is also related to the emergent literature on stochastic mortality models, following the original work by Lee and Carter (2002) and the extension by Cairns, Blake, and Dowd (2006), in that our model, which we present in the next section, assumes a DIA payout yield that is stochastic and mean reverting to a time-dependent long-term “actuarial” curve. The stochastic payout yield that we model, although motivated by the future uncertainty in pricing discount rates, can also be interpreted as uncertainty around mortality rates and insurance pricing factors. We blend them into one observable state variable: the APY.

Our article is unabashedly (narrowly) normative. We offer practical guidance to individuals who would like to spend a given sum of money over a given time frame to purchase a guaranteed deferred lifetime income, but who are unsure of the best strategy and time to do so. We do not seek to explain consumer behavior, estimate consumer preferences, or make large-scale policy recommendations about payout annuity design to help increase their appeal.⁴

To give a preview of our main results, under our model, we confirm practitioners’ intuition that one should wait to purchase until payout yields revert to (close to) normal for someone who is solely concerned with maximizing expected retirement income at some fixed age. Using the canonical example we posed earlier and common pricing assumptions implied from annuity prices, a 68-year-old male who is currently offered a payout yield of 64.60% for a DIA should wait until the payout yield “hits” 120%. Note that this is almost twice the current (age-68) payout yield and a number that is expected to be reached in 2–6 years, that is, when the individual will be between age 70 and 76, depending on assumptions about the speed of reversion of pricing discount rates presented in Section 5. Technically speaking, this is the expectation of the first hitting time of the boundary of the continuation region, which is another quantity we derive in the body of the article.

The above analysis assumes the individual is risk neutral. Qualitative behavior is quite different for risk-averse consumers. Because a risk-averse buyer is concerned about the variability of his/her income at retirement and not only his/her expected level of income—which is obvious to anyone who lives in the mean-variance world of finance—the optimal policy is more subtle. The continuation region in which the consumer does not purchase any more DIAs and continues to hold the money in cash is both time- and wealth-dependent. DIAs are purchased by risk-averse individuals when payout yields increase beyond some threshold level, but the buying process is then suspended if payout yields decline. Buying resumes as rates increase until the entire (pre-allocated) budget is exhausted. We explain more about

2 For additional research on the optimal purchase of income annuities within the context of the portfolio choice literature, see, for example, Campbell and Viciara (2002), Neuberger (2003), Cocco, Franciso, and Pascal (2005), Horneff *et al.* (2009), Kojien, Nijman, and Werker (2011), Hainuat and Deelstra (2014), Dushi and Webb (2004) or Konicz and Mulvey (2013). None of these papers assumes stochasticity in the payout rate over time or analyzes the timing problem.

3 For those topics, see, for example, Benartzi, Previtro, and Thaler (2011), as well as Inkman, Lopes, and Michaelides (2011), or Previtro (2014).

4 On the matter of the optimal design of mortality-contingent claims and annuities, we refer to the work by Scott, Watson, and Hu (2011), for example.

this case in the body of the article, whose solution echoes the literature on hedging contingent claims with transaction costs (Davis and Norman, 1990). In fact, this buying strategy is a type of dollar-cost averaging, which although has been long dismissed in the investment literature, might have some merit when it comes to DIA purchases.

We stress that our model and results are derived using a minimal set of preference parameters and assumptions. We model and focus on the dynamic behavior of DIA payout yields. But, we abstract from portfolio considerations (e.g., the magnitude of the equity risk premium), as well as human capital effects (e.g., the correlation between labor income shocks and interest rates). Our CRRA utility function is a simple one for which preferences are determined by a single risk-aversion parameter. The precise details of what we assume mathematically (and what we ignore) are laid out clearly in the next section. The benefit of these more restrictive assumptions is that we need three to five fewer parameters than most other papers in this literature, and we can still say something meaningful about optimal behavior.

The rest of the article is organized as follows. In Section 2, we define the optimization problem faced by the individual and present a verification lemma that we use to solve the problem under a general utility function. In Section 3, we consider the case for which the individual is risk neutral, that is, he/she wishes to maximize his/her expected deferred annuity income beginning at a fixed time in the future. We show that maximizing expected deferred annuity income is equivalent to picking one stopping time that maximizes expected payout yield. In that section, we also obtain the expected time that the individual will annuitize all his/her wealth, as well as approximate the optimal annuitization barrier for small values of the volatility of the payout yield. Then, in Section 4, we study properties of the optimal annuitization barrier for a risk-averse individual and approximate that barrier, thereby generalizing our work in Section 3. In Section 5, we present numerical results and examples for both risk-neutral and risk-averse individuals. Section 6 concludes the article.

2. Annuity Income Problem and Verification Lemma

We assume that the individual aged x has a fixed initial investment account (e.g., \$100,000) that he/she manages in order to maximize his/her expected income from this endowment at age $y > x$. We treat age y as his/her planned (exogenous) age at which to begin receiving deferred annuity income, and we assume that his/her current consumption is met by (other) exogenous income. Thus, the money in this investment account will be used only to ensure income from age y onward. For simplicity, we assume the account does not earn any interest.⁵

Before we launch into the formal model, we first present a refresher on how insurance companies calculate prices of life annuities, in general, and DIAs, in particular. This is also relevant to what we (later) label the long-term “actuarial” curve to which prices revert. We refer to any standard actuarial textbook for the law of large numbers justification (Bowers *et al.*, 1997), but in practice a life annuity is “priced” by insurance actuaries using the following formula:

$$a(x, y) = (1 + \ell) \sum_{k=y-x}^{\omega-x} \frac{{}_k p_x}{(1 + r_k)^k}. \quad (1)$$

In this expression, the symbol ${}_k p_x$ denotes the probability that the age- x individual survives until age of $x + k$, and y is the age at which the income begins. The probability ${}_k p_x$ accounts

5 Which is not far from reality, today.

for mortality improvements for the relevant cohort (currently aged x) and includes a margin for safety, or what an economist might call a “mortality risk premium.” The insurance company discounts future life-contingent cash flows via a term structure, denoted by r_k , from the income start age of $y \geq x$ until the end of the assumed mortality table at age ω , for example, age 120.

The pricing discount rate r_k and the present value factor $(1 + r_k)^k$ are not necessarily based on the risk-free (government) curve. The r_k is company specific, generally higher than the risk-free rate and includes a default-risk premium. See, for example, the recent work by Charupat, Kamstra, and Milevsky (2015) for evidence that annuity prices do not move in lockstep with government bond yields and are (more) highly correlated with 30-year mortgage rates, albeit with a lag. These results are also echoed by Kojien and Yogo (2015). As such, it would be inaccurate to model the evolution of annuity prices by using equilibrium swap rates superimposed on a stochastic mortality model.

Finally, the pure premium is multiplied by an insurance loading factor $\ell \geq 0$, which, in practice, is in the order of 2–5% and covers expenses (as well as profits and commissions) of the insurer. All together, these factors, via the expression in Equation (1), determine the market price at which the annuity is offered. Thus, for a \$100,000 premium paid at age x , the annual income beginning at age y would be $\frac{\$100,000}{a(x,y)}$. The income a given DIA premium will generate at age y —which is the primary quantity we are interested in—is a function of the inverse of the annuity’s price at the time of purchase. This inverse is called the APY.

To model the dynamic behavior of the payout yield, one could in theory model the evolution of the entire (default-risk-adjusted) yield curve r_k , as well as impose some sort of stochastic model for the evolution of survival probabilities ${}_k p_x$, and perhaps even try to capture age-dependent insurance loads l . Although this full-blown approach would present a complete description of annuity prices, it would needlessly complicate the model, and the number of parameters to estimate would be quite large. Instead, we propose a reduced-form model of the APY, which is the inverse of the annuity price. This reduced-form model captures the salient features of average APYs (representative among a dozen or so companies), including the expectation that yields will revert to higher levels, as well as the biological process of aging.

With the refresher out of the way, here is the formal model. At any time t between now and $T = y - x$, the individual may spend money from his/her investment account to purchase a DIA. For such an annuity, let a_t denote the random price at time t charged by a representative insurer company to a person ages $x + t$ at time t of 1 unit of income payable continuously for life, beginning at age y , or time T . The time-0 price, a_0 , is the continuous-time analog of the discrete-time annuity price in Equation (1). From the price process $\mathbf{a} = \{a_t : 0 \leq t \leq T\}$, we define the payout yield process $\Pi = \{\pi_t : 0 \leq t \leq T\}$ for this representative insurance company by

$$\pi_t = \frac{1}{a_t},$$

in which we write π_t for $\pi_t(x, y)$, the notation introduced in Section 1, because we assume x and y are fixed for the individual. Age x marks the beginning of the period during which the individual will buy DIAs—think of it as the problem’s starting age—and age y marks the end of that period. Age y is the age at which the individual wishes to begin his/her deferred annuity income. In our model, age y , or equivalently time $T = y - x$, is given

exogenously. To be clear, we do not solve for, or endogenize, the optimal income initiation age, as in Farhi and Panageas (2007), for example.

We model and focus on the dynamics of Π directly, which (again) corresponds with how DIA prices are quoted in practice. In the most general “No Arbitrage” sense, the DIA payout yield should satisfy the following expression:

$$\pi_t := \left(\mathbf{E}^{\mathcal{Q}} \left[\int_{T-t}^{\infty} e^{-\int_t^{t+s} (\tilde{r}_u + \tilde{\lambda}_u) du} ds \right] \right)^{-1}. \quad (2)$$

In this very general form, the expectation is with respect to some risk-adjusted (or risk-neutral) measure denoted by \mathcal{Q} , \tilde{r}_u is a multi-factor interest rate process, and $\tilde{\lambda}_u$ is a multi-factor mortality rate process properly capturing the evolution of cohort mortality for the group buying the DIA. In fact, while we are complicating things, $\tilde{\lambda}_u$ should be correlated with \tilde{r}_u . Recall that if indeed there is a hazard-rate risk premium and if longevity or mortality is priced, the premium should depend on the state of the economy and markets. As mentioned above, a bottoms-up approach would specify models for both \tilde{r}_u and $\tilde{\lambda}_u$, including their dependence structure—then use Fubini’s Theorem to move the expectation inside the first integral—and then attempt to derive an expression for the expectation in Equation (2). Needless to say, this is a mess. A needless mess. After all, we are interested in the evolution of APYs and have no need for individual models of stochastic mortality and interest rates. We are not hedging DIAs. In fact, we are not even selling them. Rather, we are simply trying to model the evolution of their prices over the next few decades using a tractable model.

Indeed, casual examination of Equation (2) suggests that when interest rates \tilde{r}_u are high then payout yields are high and vice versa. Casual empiricism suggests that long-term interest rates move from periods of under to over historical average. There are hundreds, if not thousands, of papers in the fixed-income literature documenting this relationship. So, at this point, we take a “leap of faith” and assume that the yield we are focused on (and need to model) obeys the following expression. We explain the intuition later.

$$d\pi_s = \kappa(\bar{\pi}(s) - \pi_s) ds + \sigma \pi_s dB_s, \quad \text{for } s > t, \quad \pi_t = \pi, \quad (3)$$

in which $\bar{\pi}(t)$ is motivated by Equation (3) and formally defined by:

$$\bar{\pi}(t) = \left(\int_{T-t}^{\infty} e^{-\int_t^{t+s} (\bar{r} + \lambda_x(u)) du} ds \right)^{-1}. \quad (4)$$

We call $\bar{\pi}(t)$ the actuarial curve. In Equation (3), which is our workhorse in this article, B is a standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F} = \{\mathcal{F}_t : 0 \leq t \leq T\}, \mathbf{P})$. The parameters κ , \bar{r} , and σ are positive, and $\lambda_x(t)$ is a deterministic function of time, which one can now think of as the pricing hazard rate, or force of mortality, at time t for an individual who is age x at time 0.

Our assumed model for Π is predicated on the theory that, in the long run, the observed (market) payout yield will converge—not to a constant, but rather—to the time-dependent actuarial curve $\bar{\pi}(t)$, which decreases exponentially over time. This central function $\bar{\pi}(t)$ is what payout yields would actually look like in the absence of any random perturbations caused by fluctuating interest or mortality rates. One can perhaps call $\bar{\pi}(t)$ the “textbook curve.” In a (fictional) world in which interest and mortality rates never change, annuities would be priced and sold based on the actuarial curve $\bar{\pi}(t)$.

In reality, market payout yields are far from the actuarial utopia and the speed at which π_t reverts to $\bar{\pi}(t)$ is driven by the parameter κ in Equation (3). We should emphasize right

now and very clearly that κ has little relation to the well-known speed of reversion in traditional short-rate (e.g., Vasicek) models. The κ we are considering is for the payout yield π_t itself, which blends the uncertainty of interest rates and mortality rates, as well as company profit factors and perhaps even the cyclicity of insurance loading. It is a catch-all parameter. It can be calibrated to historical values (assuming such data were widely available). Practically speaking, it becomes a somewhat subjective parameter (or lever) in our model. Individuals who anticipate payout yields to return to normal levels very soon will assume a higher value of κ , relative to those who expect the current pricing environment to prevail for longer. We make no empirical claim to the magnitude of κ in this article. Our article (i) derives a model for optimal annuitization using our proposed process, and (ii) then selects some reasonable values for κ —casually based on a very small data set of values—to illustrate how the model would work in practice. Future research by the authors (or others) might focus on how well our model fits the data, once this sort of data becomes available. Recall that to properly pin down mean reversion, one requires very long periods of historical data (decades).

To better understand the behavior of and rationale for $\bar{\pi}(t)$, assume that the force of mortality $\lambda_x(t) \equiv \lambda$ is constant over time, so that Equation (4) implies that the long-run payout yield function is

$$\bar{\pi}(t) = (\bar{r} + \lambda)e^{(\bar{r} + \lambda)(T-t)}.$$

When $t = T$, it is intuitively clear that $\bar{\pi}(T) = \bar{r} + \lambda$ because the (immediate) annuity's payout yield is the sum of the discount rate and the mortality rate. In fact, it should be quite obvious that if pricing rates mean revert, then so would $\bar{\pi}(T)$, which only involves an added constant. When $t < T$, the payout yield is larger than $\bar{r} + \lambda$ because income is deferred, and the payout yield reflects the credit one receives by paying now and receiving income later. Moreover, the further the individual is from time T , the higher the yield $\bar{\pi}(t)$ as a result of both the reduced probability of reaching time T and the reduced present value of the income stream. So, under a constant force of mortality, the payout yield $\bar{\pi}(t)$ declines exponentially over time at a rate of $(\bar{r} + \lambda)$ until it equals $\bar{r} + \lambda$ at time T .

When the force of mortality $\lambda_x(t)$ increases with time, the expression for $\bar{\pi}(t)$ is more complex, but it, too, declines exponentially over time to the payout yield of an immediate annuity at time T . In most of what follows and specifically in the numerical results presented in Section 5, we assume that $\lambda_x(t)$ obeys a Gompertz law of mortality. Specifically,

$$\lambda_x(t) = \frac{1}{b}e^{(x+t-m)/b},$$

in which the parameter m denotes a modal value of the remaining lifetime random variable (in years), and b is a measure of dispersion (in years). More generally, one can think of m as a location parameter and b as a scale parameter. The Gompertz mortality model is standard in the insurance literature and has been used by a number of authors in finance and insurance, including the above referenced papers on the optimal timing of annuitization.

In sum, our basic assumption about the stochastic (market) payout yield is that it will converge to $\bar{\pi}(t)$ over time. This is inspired and motivated—although not necessarily derived—by a mean-reverting process for interest rates in which the short rate converges to a pre-specified time-dependent curve.

Let A_t denote the (cumulative) amount of income from deferred life annuities in force at time $t \in [0, T]$ to be payable beginning at time T . A strategy for purchasing DIAs

$\mathbf{A} = \{A_t : 0 \leq t \leq T\}$ is admissible if it is F-progressively measurable, non-negative, and non-decreasing and if \mathbf{A} is such that $W_t \geq 0$ for all $t \in [0, T]$ with probability 1. We require that \mathbf{A} be non-decreasing because once the individual purchases a DIA, he/she cannot reverse the purchase. Thus, wealth follows the dynamics

$$\begin{cases} dW_s = -\frac{1}{\pi_s} dA_s, & \text{for } s > t, \\ W_t = w, \end{cases} \tag{5}$$

in which the payout yield follows the process in Equation (3).

We assume that the individual seeks to maximize expected utility of annuity income at time T , by optimizing over admissible controls \mathbf{A} , for an increasing, (weakly) concave utility function u . Thus, we define the value function by

$$\phi(w, A, \pi, t) = \sup_{\mathbf{A}} E^{w, A, \pi, t} (u(A_T)), \tag{6}$$

in which $E^{w, A, \pi, t}$ denotes conditional expectation given $W_t = w$, $A_t = A$, and $\pi_t = \pi$. Because the individual seeks to maximize his/her expected utility of annuity income at time T , he/she will spend all his/her money to buy DIAs. In particular, if $W_{T-} > 0$ at time $T-$, then he/she will immediately spend all his/her remaining wealth W_{T-} to buy annuity income. We remind the reader that W_0 is not the individual’s entire wealth at time 0 (or age x), but rather the wealth that the individual has set aside for purchasing DIAs during the ensuing T years.

The economic justification for Equation (6) is that the amount of annuity income accumulated by time T , denoted by A_T , induces a lifetime utility of consumption, beginning at age y , given by

$$\int_0^\infty e^{-\rho t} {}_t p_y^S u(A_T) dt = u(A_T) \int_0^\infty e^{-\rho t} {}_t p_y^S dt,$$

in which we assume the individual consumes all of his/her annuity income. Here, ρ is a subjective discount rate, ${}_t p_y^S$ is the subjective survival probability, and we assume time-separable utility of consumption; see Yaari (1965) as the classic reference for this expression. Thus, once the individual reaches time T , this lifetime utility from age T onward is effectively $u(A_T)$ times a constant that is independent of the individual’s utility function. In fact, when $\rho = 0$, that constant is the individual’s expected remaining lifetime as of age y .

We present a verification lemma that we use to solve the optimization problem embodied in Equation (6). We state the verification lemma without proof because its proof is standard in the financial and insurance mathematics literature; see, for example, Wang and Young (2012). First, define the differential operator L on an appropriate set of functions by

$$Lg = g_t + \kappa(\bar{\pi}(t) - \pi) g_\pi + \frac{1}{2} \sigma^2 \pi^2 g_{\pi\pi}. \tag{7}$$

Lemma 1. Let $\Phi = \Phi(w, A, \pi, t)$ be the classical solution of the following variational inequality on $D := \mathbf{R}^+ \times \mathbf{R}^+ \times \mathbf{R}^+ \times [0, T]$.

$$\max(L\Phi, \pi\Phi_A - \Phi_w) = 0, \tag{8}$$

with terminal condition $\Phi(w, A, \pi, T) = u(A + w\pi)$ and boundary condition $\Phi(0, A, \pi, t) = u(A)$. Then, on D ,

$$\phi = \Phi.$$

Furthermore, define the continuation region C by

$$C = \{(w, A, \pi, t) \in D : \pi\phi_A(w, A, \pi, t) < \phi_w(w, A, \pi, t)\};$$

then, the optimal strategy for annuitizing one’s wealth is to purchase just enough deferred annuity income to keep (w_t, A_t, π_t, t) in \bar{C} .

Remark 1. We interpret the inequality $\pi\phi_A \leq \phi_w$ economically and provide some intuition as follows: $\pi\phi_A$ is the marginal benefit of increasing the deferred annuity income, and ϕ_w is the marginal cost of doing so. Within the continuation region C , the marginal benefit of increasing the deferred annuity income is strictly less than the cost; therefore, in the continuation region, the individual “continues” waiting and buys annuity income later. If (w, A, π, t) is not in the closure of C , then he/she will immediately buy just enough of his/her wealth Δw so that $(w - \Delta w, A + \Delta w\pi, \pi, t)$ lies on the boundary of C . Thereafter, he/she will buy just enough to stay within the closure of the continuation region \bar{C} .

Remark 2. In numerical experiments in Section 5, we compute the continuation region when the individual’s risk preferences exhibit constant relative risk aversion γ . Specifically, we set

$$u(A) = \frac{A^{1-\gamma}}{1-\gamma}, \tag{9}$$

for some $\gamma \geq 0$ with $\gamma \neq 1$. When $\gamma = 0$, the individual is said to be risk neutral. For power utility, we can reduce the dimension of the problem. Indeed, for $A > 0$, define

$$z = \frac{w}{A},$$

and define the function g on $\mathbf{R}^+ \times \mathbf{R}^+ \times [0, T]$ by

$$g(z, \pi, t) = (1 - \gamma)\phi(z, 1, \pi, t).$$

If we determine g , then we can recover ϕ on D , when $A > 0$, by

$$\phi(w, A, \pi, t) = \frac{A^{1-\gamma}}{1-\gamma}g(w/A, \pi, t). \tag{10}$$

The variational inequality in Equation (8) becomes

$$\max\left((1 - \gamma)Lg, \pi g - \frac{1 + \pi z}{1 - \gamma}g_z\right) = 0, \tag{11}$$

with terminal condition $g(z, \pi, T) = (1 + \pi z)^{1-\gamma}$ and boundary condition $g(0, \pi, t) = 1$, and the continuation region C becomes

$$C = \left\{ (w, A, \pi, t) \in D : \pi g(w/A, \pi, t) < \frac{1 + \pi z}{1 - \gamma}g_z(w/A, \pi, t) \right\}. \tag{12}$$

Remark 3. Relative risk aversion of 1 is not covered by the power utility function in Equation (9). Instead, relative risk aversion of 1 is embodied by logarithmic utility.

$$u(A) = \ln A. \quad (13)$$

For logarithmic utility, we can also reduce the dimension of the problem, as in Remark 2. Indeed, for $A > 0$, define $z = w/A$, as before, and define the function g on $\mathbf{R}^+ \times \mathbf{R}^+ \times [0, T]$ by

$$g(z, \pi, t) = \phi(z, 1, \pi, t).$$

If we determine g , then we can recover ϕ on D , when $A > 0$, by

$$\phi(w, A, \pi, t) = \ln A + g(w/A, \pi, t). \quad (14)$$

The variational inequality in Equation (8) becomes

$$\max(Lg, \pi - (1 + \pi z)g_z) = 0, \quad (15)$$

with terminal condition $g(z, \pi, T) = \ln(1 + \pi z)$ and boundary condition $g(0, \pi, t) = 0$, and the continuation region C becomes

$$C = \{(w, A, \pi, t) \in D : \pi < (1 + \pi z)g_z(w/A, \pi, t)\}.$$

Remark 4. Economic reasoning tells us that if the current payout yield is large, then the individual will spend some (and perhaps all) of his/her wealth on a DIA. On the other hand, if the current payout yield is small, then the individual will wait to buy a DIA later. Thus, under constant relative risk aversion, the continuation region C is of the form

$$C = \{(w, A, \pi, t) \in D : \pi < \pi^*(w/A, t)\},$$

for some function $\pi = \pi^*(z, t)$ that is decreasing with respect to z . If the current payout yield $\pi > \pi^*(z, t)$, with $z = w/A$, then the individual will spend just enough wealth Δw to purchase additional DIA income of $\Delta A = \pi \Delta w$ so that $\pi = \pi^*(z', t)$, in which

$$z' = \frac{w - \Delta w}{A + \pi \Delta w} < z.$$

Thereafter, the individual will employ instantaneous control to keep his/her wealth-to-annuity-income ratio z such that $\pi \leq \pi^*(z, t)$. In Section 4, we approximate the barrier $\pi = \pi^*(z, t)$ to order σ^2 , and in numerical experiments in Section 5, we graphically display the barrier. Furthermore, when the individual is risk neutral, we show in Section 3 that the barrier $\pi = \pi^*(z, t)$ is independent of z ; write it as $\pi = \pi^*(t)$. In this case, if the current payout yield $\pi \geq \pi^*(t)$, then the individual will spend all of his/her wealth on a DIA; otherwise, the individual will wait until π_t reaches π^* .

3 Risk-neutral Purchaser

Suppose $u(A) = A$ in Equation (6), or equivalently, $\gamma = 0$ in Equation (9); then, the individual seeks to maximize the expected value of his/her annuity income at time T . In this case, the optimization problem in Equation (6) is linear with respect to the wealth of the individual. It follows that if it is optimal to spend any of the wealth in one's investment account on a DIA, then it will be optimal to spend all of the wealth in that account on a DIA. Thus, the problem becomes one of optimal stopping in that, if $W_{t-} > 0$, then there is a single time τ_t^*

in $[t, T]$ at which the individual will spend all of his/her wealth on a DIA. We have the following proposition that follows directly from Lemma 1 for this case.

Proposition 1. *Suppose $u(A) = A$ in Equation (6). Let $f = f(\pi, t)$ be the classical solution of the following variational inequality on $\mathbf{R}^+ \times [0, T]$.*

$$\max(Lf, \pi - f) = 0, \tag{16}$$

with terminal condition $f(\pi, T) = \pi$. Then, on $D = \mathbf{R}^+ \times \mathbf{R}^+ \times \mathbf{R}^+ \times [0, T]$,

$$\phi = A + wf.$$

The continuation region C is given by

$$C = \{(w, A, \pi, t) \in D : f(\pi, t) > \pi\};$$

and if the individual has not annuitized his/her wealth at time t , then the optimal time to annuitize his/her wealth is

$$\tau_t^* = \inf \{s \geq t : f(\pi, s) = \pi\}.$$

Remark 5. In the setting of Remark 2, f and g are related by $g(z, \pi, t) = 1 + zf(\pi, t)$, with $\gamma = 0$.

Remark 6. Essentially, Proposition 1 tells us that maximizing one’s expected annuity income at time T is equivalent to maximizing one’s expected payout yield between now and time T (Peskir and Shiryaev, 2006). Specifically, we can rewrite f as follows.

$$f(\pi, t) = \sup_{\tau \in \mathcal{S}[t, T]} \mathbf{E}^{\pi, t}(\pi_\tau), \tag{17}$$

in which $\mathcal{S}[t, T]$ is the set of stopping times with values in $[t, T]$.

We have the following proposition that gives us the form of the continuation region and corresponding optimal stopping time.

Proposition 2. *There exists a function of time $\pi = \pi^*(t)$, with $\pi^*(T) = 0$, such that the continuation region has the form*

$$C = \{(w, A, \pi, t) \in D : 0 \leq \pi < \pi^*(t)\},$$

in which $\pi^(t) \geq \bar{\pi}(t)$ for $0 \leq t < T$. Thus, if the individual has not annuitized his/her wealth at time $t-$, then the individual will do so at time*

$$\tau_t^* = \inf \{s \geq t : \pi_s \geq \pi^*(s)\}. \tag{18}$$

Proof. Note that $L\pi = \kappa(\bar{\pi}(t) - \pi) > 0$ for all $\pi < \bar{\pi}(t)$; thus, $f(\pi, t) > \pi$ for all $\pi < \bar{\pi}(t)$, from which it follows that the continuation region contains the set $\{(w, A, \pi, t) \in D : 0 \leq \pi < \bar{\pi}(t)\}$.

Suppose there is an interval of the form $(a(t), b(t))$ in the time- t , π -cross-section of the continuation region, with $0 < a(t) < b(t) < \infty$, whose closure is disjoint from the remainder of the time- t , π -cross-section of C . The function π is linear on $[a(t), b(t)]$; thus, π is the largest among convex functions $\psi(\pi)$ on $[a(t), b(t)]$ with $\psi(a(t)) = a(t)$ and $\psi(b(t)) = b(t)$. We also know that f is convex with respect to π because it is the supremum of convex functions. Also, $f(a(t), t) = a(t)$ and $f(b(t), t) = b(t)$ because $a(t)$ and $b(t)$ are not in the time- t ,

π -cross-section of C . It follows that $\pi \geq f(\pi, t)$ on $(a(t), b(t))$, which contradicts the fact that $f(\pi, t) > \pi$ on the continuation region. Thus, no such disjoint interval exists in the time- t , π -cross-section of the continuation region, and the proposition follows. \square

Remark 7. It is intuitively pleasing that the optimal strategy is to buy a DIA when its payout yield π is large enough, or equivalently, when the price of the DIA is low enough. See Dixit and Pindyck (1994, Appendix B of chapter 4) for an accessible discussion of threshold strategies for optimal stopping problems. They provide a rule of thumb for determining optimal thresholds, which in our case translates to the following: Because $L\pi = \kappa(\bar{\pi}(t) - \pi)$ is decreasing with respect to π , the optimal purchasing strategy is to purchase a DIA when π is large enough.

Remark 8. Note that the optimal exercise boundary $\pi = \pi^*(t)$ is not continuous at $t = T$ because we are effectively forcing the individual to buy an annuity on or before time T . Indeed, $\pi^*(T) = 0$ but $\bar{\pi}(T) > 0$, and Proposition 2 tells us that $\pi^*(t) \geq \bar{\pi}(t)$ for all $0 \leq t < T$. Alternatively, we could define $\pi^*(T) = \lim_{t \rightarrow T^-} \pi^*(t)$ and say that the individual buys annuities at $\tau_t^* \wedge T$, in which τ_t^* is given in Equation (18).

As σ increases, the payout yield's process becomes more volatile, and the maximum payout yield has a greater probability of becoming larger; therefore, we expect the value function f and the optimal exercise boundary to increase with σ , which we prove in the next proposition.

Proposition 3. *For a risk-neutral individual, the value function ϕ defined in Equation (6) and the optimal exercise boundary π^* from Proposition 2 are non-decreasing with respect to σ .*

Proof. Suppose $\sigma_1 < \sigma_2$. Let f^i and π_i^* denote the function given in Proposition 1 and optimal exercise boundary, respectively, when $\sigma = \sigma_i$ for $i = 1, 2$. Define $F = f^2 - f^1$. For $\pi \geq \pi_1^*(t)$, we have $f^1(\pi, t) = \pi$, so $F(\pi, t) = f^2(\pi, t) - \pi \geq 0$.

For $\pi < \pi_1^*(t)$, f^1 satisfies $L^1 f^1 = 0$ and f^2 satisfies $L^2 f^2 \leq 0$, in which L^i is given by Equation (7) with $\sigma = \sigma_i$ for $i = 1, 2$. Thus, $L^2 f^2 - L^1 f^1 \leq 0$, or equivalently

$$L^2 F \leq -\frac{1}{2} \pi^2 (\sigma_2^2 - \sigma_1^2) f_{\pi\pi}^1,$$

and the right side of this inequality is non-positive because $f_{\pi\pi}^1 \geq 0$. Recall that f^1 is convex, which we showed in the proof of Proposition 2.

To summarize, we have $F \geq 0$ for $\pi \geq \pi_1^*(t)$, $L^2 F \leq 0$ for $\pi < \pi_1^*(t)$, and $F(\pi, 0) = 0$ for $\pi \in \mathbf{R}^+$. It follows from the maximum–minimum principle (Walter, 1970, section 26) that $F \geq 0$ or equivalently that $f^1 \leq f^2$ on $\mathbf{R}^+ \times [0, T]$, which implies that ϕ is non-decreasing with respect to σ because $\phi = A + wf$.

Next, we show that $\pi_1^* \leq \pi_2^*$ on $[0, T]$. Suppose that this inequality does not hold for some $t_0 \in [0, T]$; then, let $\pi_0 \in (\pi_2^*(t_0), \pi_1^*(t_0))$. We have $f^2(\pi_0, t_0) = \pi_0$ and $f^1(\pi_0, t_0) > \pi_0$, which contradicts $f^1 \leq f^2$ on $\mathbf{R}^+ \times [0, T]$. Thus, no such π_0 and t_0 exist, and we have shown that $\pi_1^* \leq \pi_2^*$ on $[0, T]$. \square

Proposition 3 resonates with the option pricing literature, for example, where it is well known that the value of an American put increases with the volatility of the underlying stock (Brennan and Schwartz, 1977). In numerical experiments not shown here, we solved the variational inequality in Equation (16) and observed that f and π^* increase with respect to σ .

Given the optimal exercise boundary $\pi^*(t)$ for our risk-neutral individual to purchase a DIA, it is natural to ask, “At time t , how much longer will the individual have to wait to purchase a DIA (on average)?” We next answer this question. If the individual has not annuitized his/her wealth before time t , then the expected waiting time is the expected value of $\tau_t^w = \tau_t^* - t$. We use a superscript w to indicate that τ_t^w is the waiting time until the payout yield reaches the optimal exercise boundary π^* . Recall that because we defined $\pi^*(T) = 0$, if $W_{T-} > 0$, then $\tau_T^w = 0$.

Define the expected waiting time h by

$$h(\pi, t) = \mathbf{E}^{\pi, t}(\tau_t^w); \tag{19}$$

then, h is the unique classical solution of the following boundary-value problem for $(\pi, t) \in [0, \pi^*(t)] \times [0, T]$.

$$\begin{cases} 1 + h_t + \kappa(\bar{\pi}(t) - \pi) h_\pi + \frac{1}{2} \sigma^2 \pi^2 h_{\pi\pi} = 0, \\ h(\pi^*(t), t) = 0, \\ h(\pi, T) = 0. \end{cases} \tag{20}$$

Furthermore, for $\pi > \pi^*(t)$, we have $h(\pi, t) = 0$. As σ increases, we know from Proposition 3 that π^* increases; thus, we expect the time until hitting that boundary to increase with σ , also. In numerical experiments not shown here, we computed the expected waiting time and observed that it increases with respect to σ .

We present a σ^2 -order approximation of the free boundary $\pi^*(t)$ in the next proposition; see Appendix A.1 for its proof.

Proposition 4. *To order σ^2 , the free boundary π^* is approximated by $\hat{\pi}$, in which*

$$\hat{\pi}(t) = \bar{\pi}(t) \left(1 + \frac{\sigma^2}{2} \frac{1}{\bar{r} + \lambda_x(t)} \right), \tag{21}$$

for $0 \leq t < T$.

Remark 9. Note that the approximation of $\pi^*(t)$ in Equation (21) is greater than $\bar{\pi}(t)$ for $0 \leq t < T$, as expected from Proposition 2. Also, the approximation increases with respect to σ , as expected from Proposition 3. In Appendix C, we observe that the approximation in Equation (21) is excellent when σ is small, specifically $< 5\%$, which equals the average volatility for the companies selling 20-year deferred DIAs to individuals aged 68; see Section 5.

Proposition 4 leads to the question “If the individual were to follow the purchasing strategy dictated by the approximation in Equation (21), that is, if he/she were to spend all his/her money on annuity income as soon as $\pi \geq \hat{\pi}$, then how closely would the resulting value function \bar{f} approximate the optimal value function f (or $\phi = A + wf$)?” Proposition 5 below answers this question; see Appendix A.2 for its proof. As an aside, we cannot use the notation \hat{f} for the approximate value function because \hat{f} is used for a different approximation of f in Appendix A.1.

Define the time that the payout yield reaches $\hat{\pi}$ by

$$\hat{\tau}_t = \inf \{s \geq t : \pi_s \geq \hat{\pi}(s)\}. \tag{22}$$

which is analogous to the expression for τ_t^* in Equation (18). Define the corresponding expected value of the payout yield when it reaches $\hat{\pi}$ by

$$\bar{f}(\pi, t) = \mathbf{E}^{\pi, t}(\pi_{\tau_t}),$$

which is analogous to $f(\pi, t) = \mathbf{E}^{\pi, t}(\pi_{\tau_t^*})$. Then, \bar{f} is the unique classical solution of the following boundary-value problem for $(\pi, t) \in [0, \hat{\pi}(t)] \times [0, T]$.

$$\begin{cases} \bar{f}_t + \kappa(\bar{\pi}(t) - \pi)\bar{f}_\pi + \frac{1}{2}\sigma^2\pi^2\bar{f}_{\pi\pi} = 0, \\ \bar{f}(\hat{\pi}(t), t) = \hat{\pi}(t), \\ \bar{f}(\pi, T) = \pi. \end{cases} \tag{23}$$

For $\pi > \hat{\pi}(t)$, we have $\bar{f}(\pi, t) = \pi$.

Proposition 5. There exists a positive constant β such that

$$\bar{f} \leq f \leq \bar{f} + \beta\sigma^2,$$

uniformly over all $(\pi, t) \in \mathbb{R}^+ \times [0, T]$.

Remark 10. Proposition 5 tells us that the strategy determined by the approximation $\hat{\pi}$ results in an error of order $O(\sigma^2)$ for the value function, not $O(\sigma^4)$, as one might expect. It also tells us that additional terms in the approximation will not necessarily lead to a more accurate approximate value function, which conforms with our numerical work (data not shown in this article).

4. Risk-averse Purchaser

In this section, we assume that preferences exhibit constant relative risk aversion $\gamma > 0$, and we focus on the case for which $\gamma \neq 1$. Recall that if we define g by

$$g(z, \pi, t) = (1 - \gamma)\phi(z, 1, \pi, t),$$

then the value function ϕ is given by

$$\phi(w, A, \pi, t) = \frac{A^{1-\gamma}}{1-\gamma}g(w/A, \pi, t),$$

when $A > 0$. From Remark 2 and from the relationship between variational inequalities and free-boundary problems, on the closure of the continuation region, g solves the following free-boundary problem.

$$\begin{cases} g_t + \kappa(\bar{\pi}(t) - \pi)g_\pi + \frac{1}{2}\sigma^2\pi^2g_{\pi\pi} = 0, & 0 \leq \pi \leq \pi^*(z, t), \\ \frac{\partial}{\partial z} \ln g(z, \pi^*(z, t), t) = (1 - \gamma) \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}, \\ \frac{\partial}{\partial z} \ln g_z(z, \pi^*(z, t), t) = -\gamma \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}, \\ g(0, \pi, t) = 1, \\ g(z, \pi, T) = (1 + \pi z)^{1-\gamma}, & 0 \leq \pi < \pi^*(z, T-). \end{cases} \tag{24}$$

The higher order smooth fit condition $\frac{\partial}{\partial z} \ln g_z(z, \pi^*(z, t), t) = -\gamma \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}$ comes from the optimality of the free boundary (see, Dixit, 1991 or Dumas, 1991); specifically, it comes from $\pi \phi_{Aw} = \phi_{ww}$ at the free boundary.

As the risk-averse buyer spends more of his/her wealth on DIAs and locks in his/her time- T annuity income, we expect that, as wealth approaches zero, he/she will be willing to wait to buy annuities until the payout yield reaches the boundary for the risk-neutral buyer. We prove this in the following proposition.

Proposition 6. *The free-boundary for the risk-averse buyer approaches the free-boundary for the risk-neutral buyer as wealth approaches zero, that is, $\pi^*(0, t) = \pi^*(t)$ for all $\gamma \geq 0$.*

Proof. To prove this result for $\gamma > 0$ and $\gamma \neq 1$, we show formally that $\frac{1}{1-\gamma} g_z|_{z=0}$ solves the same free-boundary problem that f solves in (Appendix A.1). Thus, by uniqueness of the solutions of free-boundary problems, we conclude that $\pi^*(0, t) = \pi^*(t)$ for all $\gamma \geq 0$ and $\gamma \neq 1$.

To this end, first observe that because $Lg = 0$ on the closure of the continuation region and because the operator L is independent of z , it follows that

$$L\left(\frac{1}{1-\gamma} g_z|_{z=0}\right) = 0.$$

Next, because $g(z, \pi, T) = (1 + \pi z)^{1-\gamma}$ for all $0 \leq \pi < \pi^*(z, T-)$, it follows that $g_z(z, \pi, T) = (1 - \gamma)\pi(1 + \pi z)^{-\gamma}$. Thus,

$$\frac{1}{1-\gamma} g_z(0, \pi, T) = \pi, \quad 0 \leq \pi < \pi^*(0, T-).$$

Next, because $\frac{\partial}{\partial z} \ln g(z, \pi^*(z, t), t) = (1 - \gamma) \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}$ and because $g(0, \pi, t) = 1$, it follows that

$$\frac{1}{1-\gamma} g_z(0, \pi^*(0, t), t) = \pi^*(0, t).$$

Finally, if we fully differentiate $\frac{\partial}{\partial z} \ln g(z, \pi^*(z, t), t) = (1 - \gamma) \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}$ with respect to z and simplify, we obtain

$$\frac{\partial}{\partial \pi} \left(\frac{1}{1-\gamma} g_z(0, \pi^*(0, t), t) \right) = 1.$$

In simplifying the full derivative, we use $g_{zz}(0, \pi^*(0, t), t) = -\gamma(1 - \gamma)(\pi^*(0, t))^2$ and $g_{\pi}(0, \pi, t) = 0$, in addition to $g(0, \pi, t) = 1$ and $g_z(0, \pi^*(0, t), t) = (1 - \gamma)\pi^*(0, t)$. Thus, $\frac{1}{1-\gamma} g_z|_{z=0}$ solves the free-boundary problem in (Appendix A.1), from which it follows that the free-boundary for the risk-averse buyer equals the free-boundary for the risk-neutral buyer when $z = 0$.

The proof for $\gamma = 1$ follows along the same lines by showing that $g_z|_{z=0}$, in which g is given in Remark 3, solves the free-boundary problem in (Appendix A.1). □

Remark 11. If the individual begins with no pre-existing deferred annuity income, that is, if $A = 0$, then he/she will begin to buy annuity income when $\pi_t \geq \lim_{z \rightarrow \infty} \pi^*(z, t)$ and will stop when π_t reaches $\pi^*(t)$, the risk-neutral exercise boundary. Thus, the so-called resident time, the expected time during which the individual purchases annuity income, is bounded above by the expected waiting time h defined in Equation (19).

We provide a σ^2 -order approximation of the free boundary $\pi^*(z, t)$ in the next proposition; see Appendix B for its proof.

Proposition 7. *To order σ^2 , the free boundary π^* is approximated by $\hat{\pi}$, in which*

$$\hat{\pi}(z, t) = \bar{\pi}(t) \left(1 + \frac{\sigma^2}{2} \frac{1}{\bar{r} + \lambda_x(t)} - \gamma \frac{\sigma^2}{\kappa} \frac{\bar{\pi}(t)z}{1 + \bar{\pi}(t)z} \right). \quad (25)$$

Remark 12. Note that the approximation in Equation (25) generalizes the one in Equation (21) for the risk-neutral purchaser. Indeed, if $\gamma = 0$ in Equation (25), then we retrieve Equation (21). Also, in work not shown in this article, we independently obtained an approximation for a risk-averse purchaser with constant relative risk aversion of $\gamma = 1$ (i.e., logarithmic utility), and it matches the approximation in Equation (25) with γ set equal to 1. In Appendix C, we observe that the approximation in Equation (25) is excellent when σ is small, specifically $<5\%$, and when $\gamma < 5$.

Remark 13. It is intuitively pleasing that the approximation in Equation (25) decreases with respect to γ because we expect a more risk-averse individual to buy DIA sooner than a less risk-averse individual. Moreover, as the individual buys more and more DIA, z decreases. The approximation in Equation (25) increases as z decreases; thus, the individual is more willing to wait for better payout yields as he/she locks more of his/her wealth into the DIA. Also, as z approaches 0, the approximation in Equation (25) approaches the one for the risk-neutral purchaser in Equation (21), so all individuals finish spending all their wealth on DIAs at the same time, regardless of risk aversion. Thus, the approximations in Equations (21) and (25) are consistent with Proposition 6.

5. Numerical Results

In this section we illustrate and apply our model. We specify the optimal purchasing strategy for a particular DIA, namely one purchased at age $x = 68$ with income beginning at age $y = 88$. According to data from CANNEX Financial Exchange, the median purchase age for (income) annuities is approximation age 68, and the $T = 20$ deferral period takes us to age $y = 88$ which is the latest age at which all companies are willing to still quote DIAs. Additional examples with different ages are addressed as well.

We estimated $\sigma(68, 88)$ for our model in Equation (3) by first computing the historical volatility of π_t for each of the (short) time series from insurance companies in our database during the period from May 2014 to January 2015. We observed individual volatilities for $\sigma(68, 88)$ ranging from 3% (for companies that rarely change their prices) to 12% (for companies that are quite responsive to changing interest rates). Then, to estimate the value of σ for a representative insurance company, we averaged those individual volatilities to get $\hat{\sigma} = 5\%$.

To get a feeling for the optimal purchasing strategy we generated a sample path for π from Equation (3) via Monte Carlo simulation, by using a rate of mean-reversion of $\kappa = 10\%$, a volatility of $\sigma = 5\%$, and an initial payout yield of $\pi_0 = 64.6\%$ for a DIA at age 68, deferred for 20 years to age 88; see Figure 2. The coefficient of relative risk aversion of the individual is a (relatively) high value of $\gamma = 10$, and the assumed long-term pricing discount rate is $\bar{r} = 5\%$. The Gompertz parameters for calculating survival probabilities ${}_t p_x$, used in $\bar{\pi}(t)$, are $m = 87.65$ (modal value in years) and $b = 11.5$ (dispersion value in

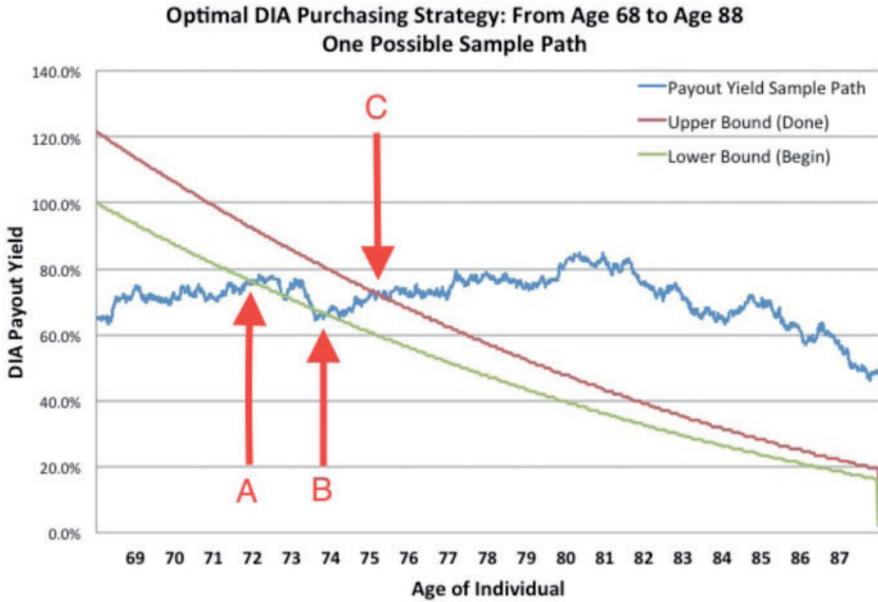


Figure 2. Following a sample path over time. Each set of values for γ, κ, σ , Gompertz parameters m, b , and the long-term pricing rate \bar{r} generates upper and lower bound curves which bracket the annuitization region.

years).⁶ These two parameters were calibrated to the earlier-mentioned quotes available in late December 2014. Later we illustrate different γ values.

We re-iterate that our value for $\kappa = 10\%$ is rather casual, and is not necessarily equal to (or even in close in value) to the common parameters used in a short-rate (e.g., Vasicek) model calibrated to term-structure derivatives. Rather, we could not reject the hypothesis value of $\kappa = 10\%$ using our limited data, and we use other estimates later.

For the sample path in Figure 2, during the first four years, from the (initial) age of 68 until age 72, the individual does not purchase any DIA units because the payout yield is expected to improve over time. Then, at age 72, the individual begins buying DIAs (point A in the figure) because the payout yield crosses the lower bound, given by $\lim_{z \rightarrow \infty} \pi^*(z, t)$, in which $z = \frac{w}{A}$. Remember that a risk-averse individual will not spend all his/her budget on DIA units at one time, as will a risk-neutral purchaser. This risk-averse individual buys DIA units occasionally and only for a few months. Then, at age 73.5, even though he/she has not yet spent his/her entire budget, he/she stops buying (point B in the figure) because the payout yield is too low. In our previous notation, the payout yield has dropped below $\pi^*(z, t)$.

The purchasing suspension, if one can call it that, lasts for about three months and then resumes because the payout yield is greater than $\pi^*(z, t)$. The amount purchased at any point within the region bounded by the two curves is governed by Lemma 1. Technically

6 These parameters result in an expected future lifetime of 17.98 years for our 68-year-old, or equivalently, an expected age at death of approximately 86 years. Recall that the modal value of life is (approximately) $m = 87.6$ years.

speaking, the payout yield crosses $\pi^*(z, t)$ infinitely many times in continuous time, which can be measured using the local time of the process. In practice, one would engage in buying once per month (or perhaps once per quarter), so the likelihood of starting and then stopping and then starting again would be diminished.

Finally, at the point labeled C, which is age 75 in this particular scenario, the sample path exits the buying region for the first time; that is, the payout yield becomes greater than $\pi^*(0, t)$. In other words, at point C, the individual has spent his/her entire budget on DIAs. Even if the sample path were to re-enter the curved region from above—which it does for a few months and infinitely often in continuous time—the individual has no more wealth to spend on DIAs. The first passage time across $\pi = \pi^*(0, t)$ is the last buying date.

Of course, the above “story” is for one particular sample path for future values of π_s . How does one derive the buying and stopping (and then buying and stopping again) strategy for a general sample path? The answer is in the following scheme.

5.1 Iterative Scheme

What we call the purchasing function, or the amount of wealth that is spent at any time period j and denoted by ΔW_j , can be computed explicitly using our analytic approximation to the free boundary, presented in Equation (25). Computationally and iteratively this is done as follows. First, let π_j denote the current (market) payout rate, and assume that we just entered the “purchase” region, so that $\pi_j > \pi_j^*$. (We use subscript j throughout this scheme to denote the time period j .) Then, define C_j by

$$C_j = \max \left\{ \left(\frac{\sigma^2}{2(\bar{r} + \lambda_j)} - \frac{\pi_j - \bar{\pi}_j}{\bar{\pi}_j} \right) \frac{\kappa}{\gamma \sigma^2}, 0 \right\}, \tag{26}$$

in which λ_j denotes the appropriate mortality rate at the relevant age and time. The value of C_j depends on all seven of our key parameters ($\kappa, \sigma, \bar{r}, \gamma, x, m, b$), and it uniquely determines the amount of (remaining) wealth ΔW_j that should be removed from cash and converted into DIAs.

In the second step of the iterative scheme, which solves for the purchasing function, we invert the key Equation (25) and define a “target” value of z , to which the portfolio consisting of cash and DIAs should be rebalanced. The target ratio, which is denoted by z' , as it was in the earlier part of the article, is given by

$$z' = \frac{C_j}{\bar{\pi}_j(1 - C_j)}. \tag{27}$$

This target ratio depends on the value of the “actuarial” curve via $\bar{\pi}_j$ and on the above-defined C_j . Recall that prior to the (new, additional) purchase, the value of $z = W_{j-1}/A_{j-1}$. After the purchase of DIAs, the ratio must equal the target value z' ; thus, to achieve this target ratio, let

$$\Delta W_j = \frac{W_{j-1} - z' A_{j-1}}{z' \pi_j + 1}. \tag{28}$$

ΔW_j is the amount of money that is spent on DIAs at time period j . Technically speaking, since the iterative scheme operates in discrete time, we must have $\Delta W_j \leq W_{j-1}$; in other words, the individual cannot spend more than his/her remaining wealth or dedicated budget. This is strictly enforced here and the equality occurs when $C_j = 0$ and $z' = 0$. In continuous time, this would never be an issue, since π_j can never get too far from π_j^* .

To close the algorithm (or iterative scheme), construct the next value of wealth by $W_j = W_{j-1} - \Delta W_j$. The loop continues with a (new) fresh value of π_{j+1} —obtained via simulation or the market itself—until wealth is driven to zero. Of course, if and when the value of π_{j+1} lies in the continuation region (i.e., is less than π_{j+1}^*), the purchasing process is halted (temporarily).

5.2 Working with the Scheme

Assume we are modeling an individual or a cohort that is (exactly) 55 years of age with a target date (or retirement) horizon of age 75. For this group, suppose the current APY in the market is quoted at $\pi_0 = 36\%$, whereas the long-run actuarial assumption model prices this DIA with a payout yield of $\bar{\pi}_0 = 39.85\%$. The long-run actuarial interest rate of $\bar{r} = 5\%$, and the time-0 mortality rate is $\lambda_0 = 0.5081\%$. Moving to the other input parameters, if the underlying volatility, reversion rate, and risk-aversion parameters are $\sigma = 5\%$, $\kappa = 10\%$, and $\gamma = 5$, respectively, then the value of C_0 is 0.9544, according to Equation (26). Assume that this individual or cohort has a total budget for annuities of \$50,000 and no prior annuity income. Using our notation, $W_0 = \$50,000$, $A_0 = \$0$, the current (market) APY $\pi_0 = 36\%$, and the target-ratio variable $z' = 52.52$ from Equation (25); then, according to Equation (28), the value of $\Delta W_0 = \$2,512$. A total of slightly over \$2,500 should be spent to purchase DIAs today.

In contrast to the above situation, assume the current DIA payout yield is 40% (which is 400 basis points higher). We keep the other six parameters at the same values. This time, $C_0 = 0.1514$ and the target ratio Equation (27), is $z' = 0.4474$ according to Equation (25). Finally, assuming \$50,000 of liquid (non-annuitized) wealth and no prior annuity income, the algorithm indicates a rather large purchase of DIAs. In this case, the optimal amount to spend is $\Delta W_0 = \$42,406$, which is nearly the entire budget allocated of \$50,000. The economic intuition for the large purchase is that current payout yields for our 55-year-old ($\pi_0 = 40\%$) is at the textbook payout yield $\bar{\pi}_0 = 39.85\%$. Hence, it is optimal to spend most—although not the entire—remaining budget on annuities, because (i) they might get more expensive later, and (ii) even if yields stay at current levels, while holding DIA units the investor is earning mortality credits.

5.3 The Strategy over Time

Figure 3 displays the result of this iterative scheme for a sample path (generated, again, via Monte Carlo). It displays the optimal DIA (for income starting at age $\gamma = 88$) purchasing strategy from age $x = 68$ with \$100,000 assuming six levels of risk aversion γ ranging from 0 to 9. These ages are higher than and in contrast to the previous example, closer to the ages envisioned by the US Treasury department for the ALDAs which begin around age 85. Using our asymptotic approximation, for someone who is risk-neutral the entire sum is converted to DIAs at age 73.25. For higher levels of risk aversion, γ , optimal purchasing is spread out over time and (only) finalized at age 73.25. Notice how purchasing occurs in an impulse-like manner as a result of the iterative scheme described above. This should not come as a surprise to those familiar with the literature on irreversible investment, but rather it is the age structure that is interesting. As in the case of Figure 2, the sample path and corresponding strategy were generated starting at $\pi_0 = 64.6\%$ using the stochastic process described in the article with $\kappa = 10\%$, $\sigma = 5\%$, and $\bar{r} = 5\%$ and using the iterative process detailed above.

Figures 4 and 6 provide some further indication of the accuracy of our asymptotic approximation for different ages and different values of κ . For example, in Figure 4 we go

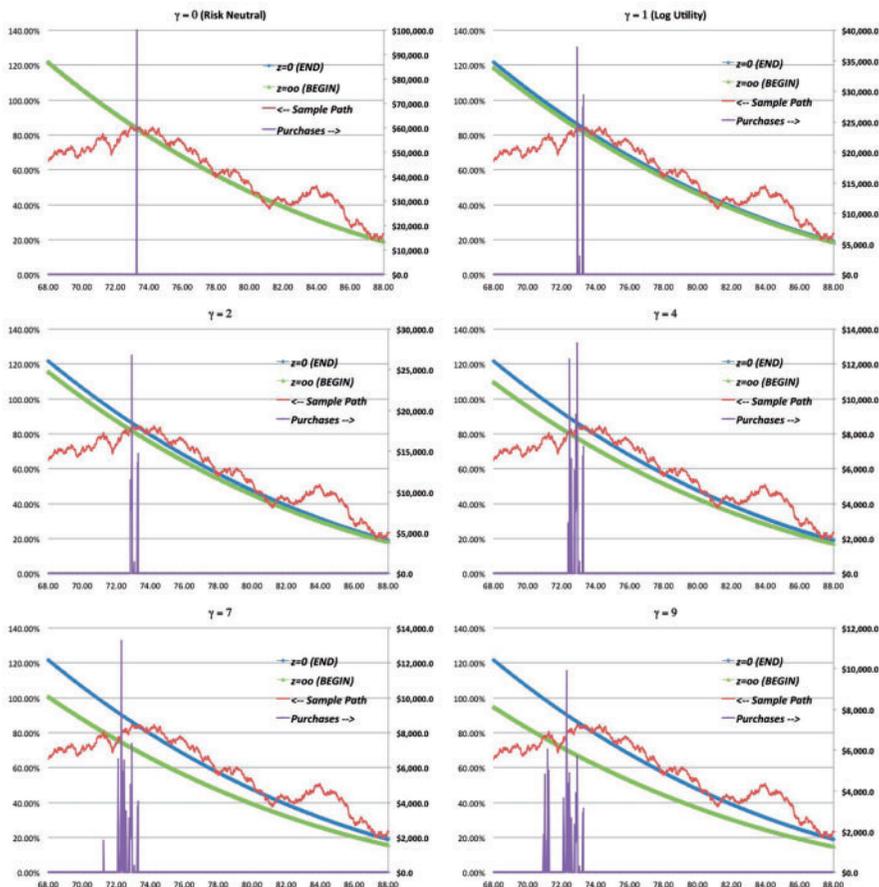


Figure 3. One sample path. Six outcomes. The above displays the optimal DIA (starting income at age $\gamma=88$) purchasing strategy from age $x=68$ with \$100,000 assuming six levels of risk aversion γ ranging from 0 to 9. Under risk-neutrality the entire sum is converted to DIAs at age 73.25. For higher levels of risk aversion γ , optimal purchasing is spread-out over time and (only) finalized at age 73.25. The (one) sample path and corresponding strategy was generated from $\pi_0 = 64.6\%$ using the stochastic process described in the article with $\kappa = 10\%$, $\sigma = 5\%$, $\bar{r} = 5\%$, and asymptotic approximation.

back to age 50 as the starting point and have the income begin at the age of 70, with a speed of mean reversion of $\kappa = 5\%$. For lower values of risk aversion $\gamma = 5$ the approximation is indistinguishable from the proper (numerical) solution of the free-boundary problem. But, when γ is increased to a value of 10, there is some deterioration in the accuracy, for lower ages. In other words, we would recommend using the numerical scheme to extract or solve for the purchasing region. Interestingly, in Figure 6, when the speed of mean reversion is doubled to $\kappa = 10\%$, the asymptotic approximation produces results that are indistinguishable from the numerical solutions, even at higher values of γ .

So, despite the rather complex optimization problem embodied in Equation (8) and the inevitability of solving that free-boundary problem numerically, this iterative procedure for approximating the optimal strategy can easily be coded in a spreadsheet. Equation (25) is the key, and using it will result in a good approximation, as long as σ and γ are small

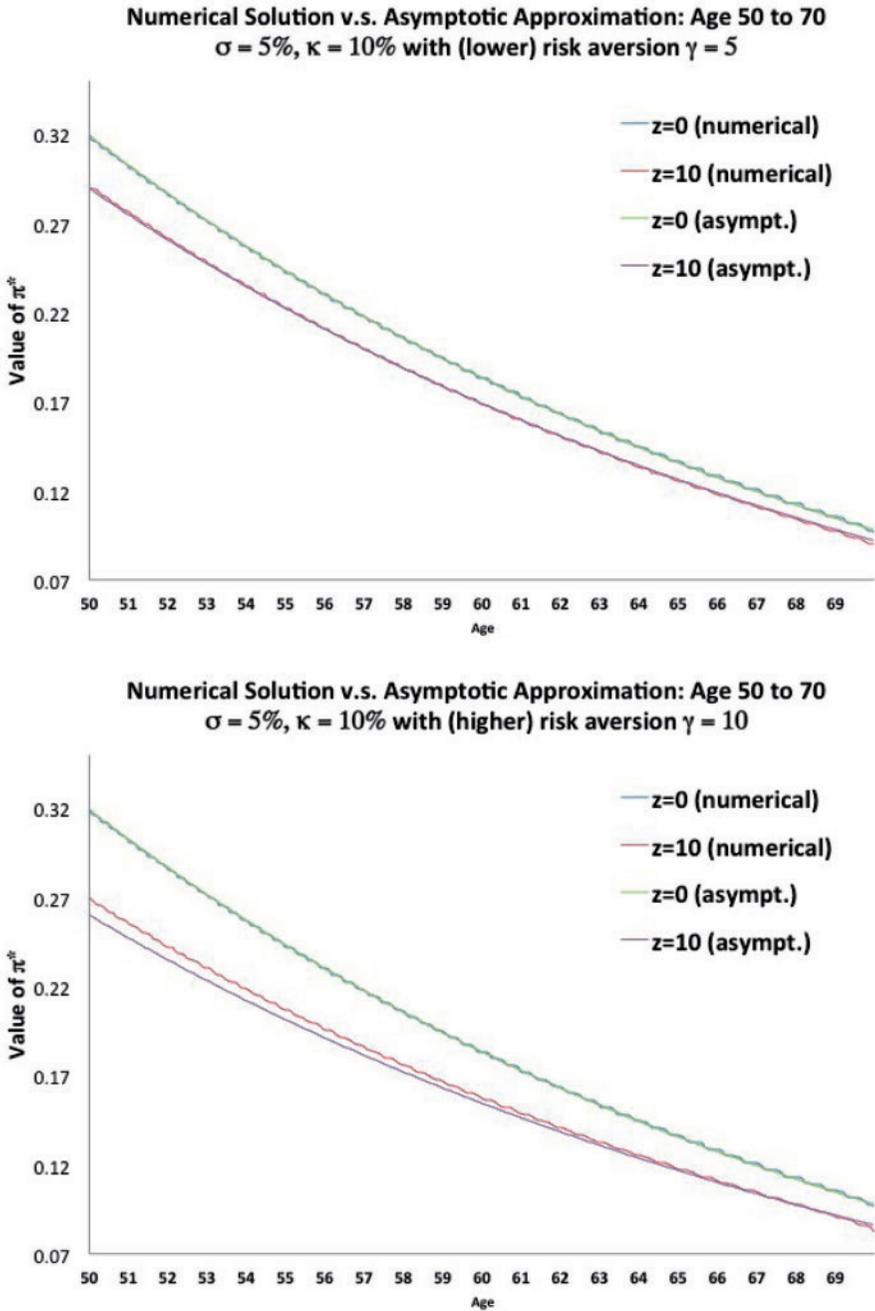


Figure 4. Analysis of approximation for DIA purchases starting at age 50 and terminating by age 70. As shown in Figure 5, for lower values of risk aversion γ (as well as σ) the asymptotic approximation of the free boundary $\pi^*(z, t)$ is quite accurate relative to the numerical solution. For higher values of γ the approximation is no worse than at older ages.

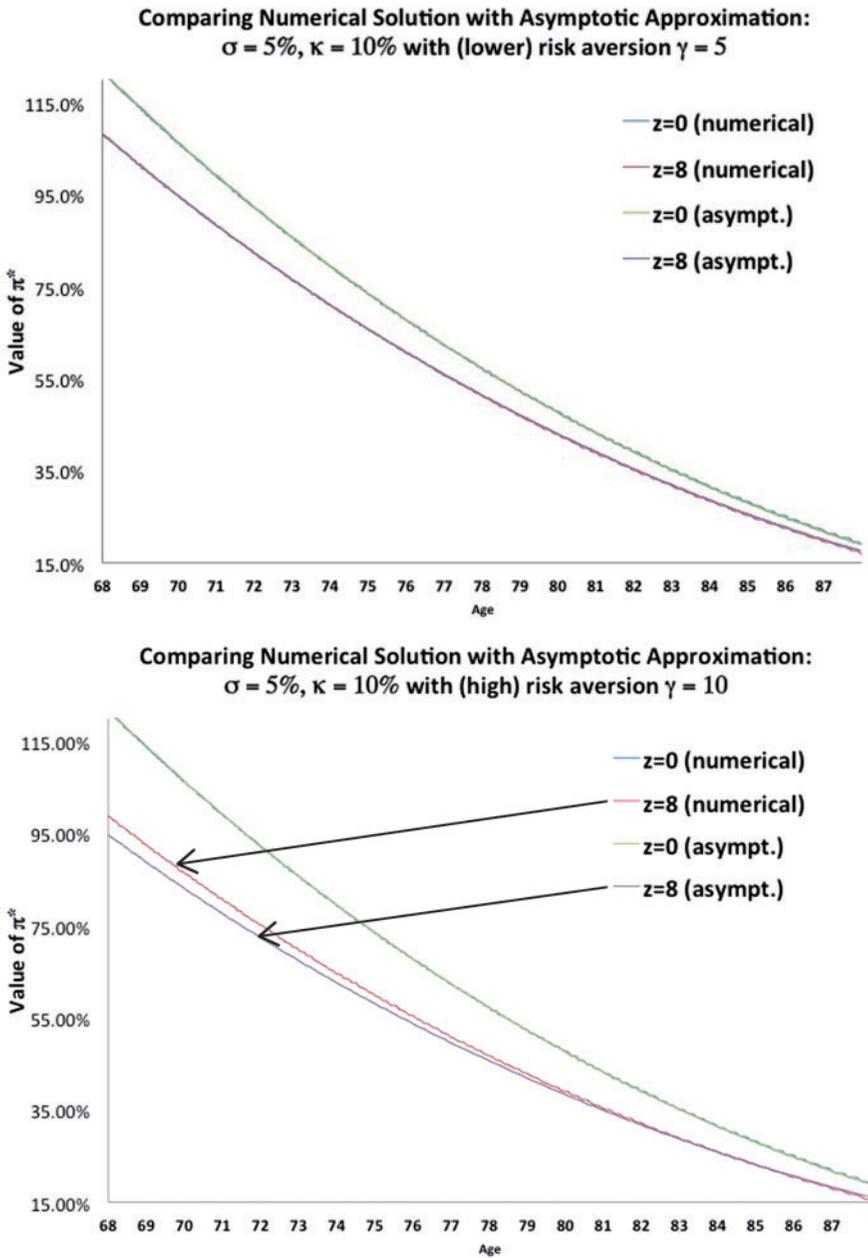


Figure 5. For lower values of risk aversion γ (as well as σ) the asymptotic approximation of the free boundary $\pi^*(z, t)$ is quite accurate relative to the numerical solution. For example, when $\gamma = 5$ the corresponding values of the two π^* curves are indistinguishable. But, for $\gamma = 10$ there is a gap for higher values of $z = W/A$. In other words, the approximation is less accurate for $z > 0$, but for $z = 0$ they (still) match.

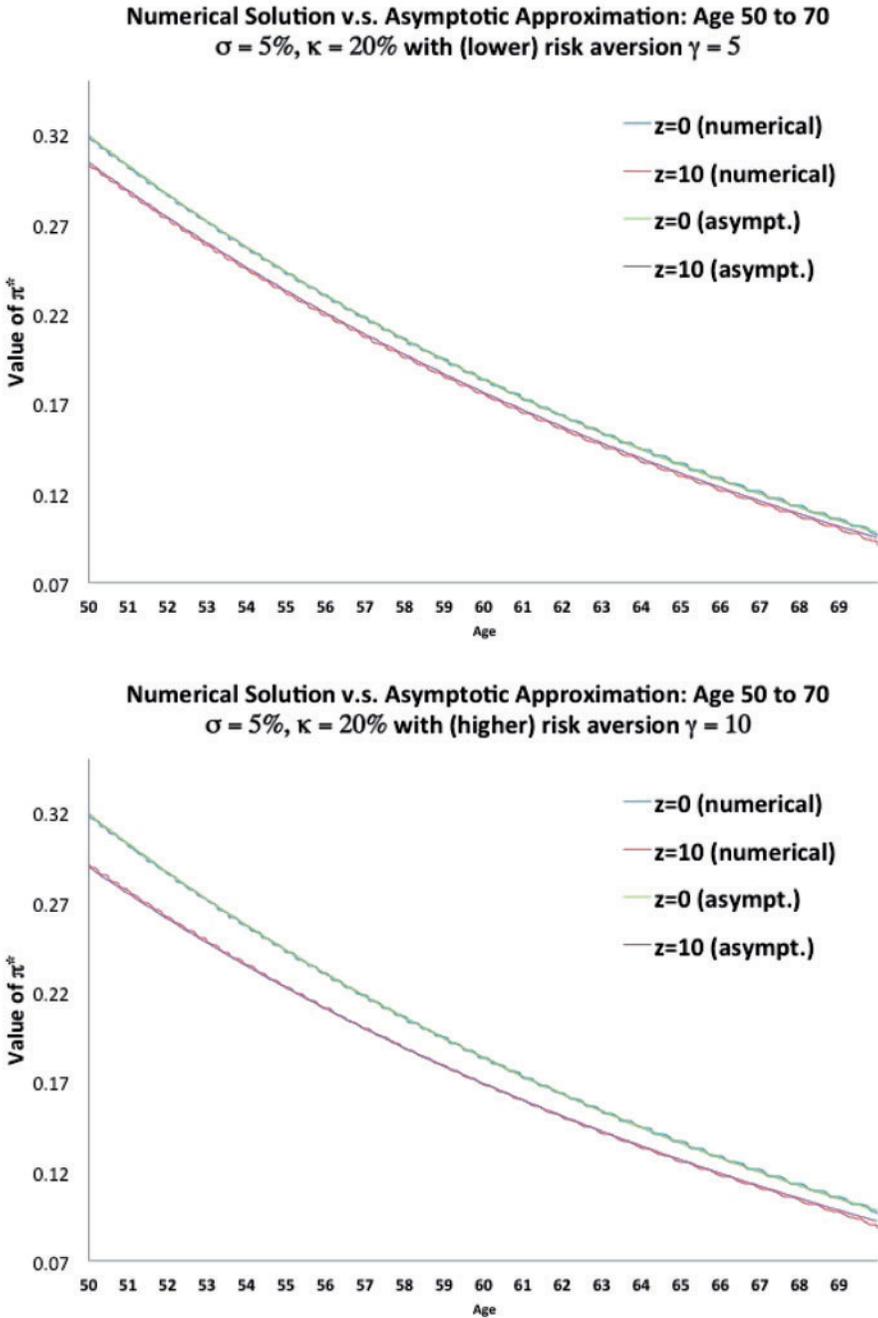


Figure 6. Analysis of approximation for DIA purchases starting at age 50 and terminating by age 70, when the value of κ is doubled from 10% to 20%. There is no noticeable difference in the accuracy of the asymptotic approximation. In other words, the approximation is relatively better for higher values of the speed of mean reversion.

enough; see Remarks 9 and 12 and Appendix C. As well, see a follow-up paper by Milevsky, Huang, and Young (2015) in which some additional numerical examples and practical insights are provided.

We emphasize that none of the above numerical examples is intended to opine on the true value of either σ or κ , nor to comment on the rather tricky problem of which value of γ to use within the context of a target-date fund or large retirement plan. All we claim is that once those parameters have been identified, we can offer a reasonable, efficient, and accurate way to purchase the DIAs optimally. We believe this to be the key contribution of the article to the finance and insurance literature.

6. Conclusion

In this article, we analyzed and determined the optimal strategy for purchasing DIAs for an individual who wishes to maximize the expected utility of his/her annuity income at a fixed time in the future. In contrast to the vast and growing portfolio-choice literature for mortality-contingent claims, we focused on the stochasticity of the DIA's payout yield and addressed concerns that payout yields are currently too low to justify irreversible annuitization. Our assumed stochastic process of payout yields is a generalized mean-reverting process.

The issue of optimal purchase of delayed annuities is of special relevance to the lifecycle and target-date investment business in light of recent announcements by both the US Treasury and Department of Labor allowing (and encouraging) the use of DIAs in 401(k) and other qualified plans.

We show that when payout yields are mean-reverting, a risk-neutral consumer who wishes to maximize his/her expected retirement income should wait until yields reach a threshold—which lies above historical averages—and then purchase the DIA in one lump sum. In contrast, we also showed that a risk-averse consumer, who is concerned payout yields will remain below average for an extended period and who worries about losing mortality credits while waiting should employ a barrier purchasing strategy. This is quite similar to optimal behavior in the portfolio problem under transaction costs but our contribution lies in illustrating how this would be applied and employed in the context of mortality-contingent claims. Furthermore we offer an accurate and easy-to-use asymptotic approximation (or scheme) for the optimal purchasing strategy and showed (in the numerical section) that compares well with numerical solutions to the associated Free Boundary Problem. Practically speaking, the optimal behavior of a risk-averse consumer resembles an asymmetric dollar-cost averaging strategy, and a portion of the “budget” for DIAs is spent even while payout rates are below historical averages.

In ongoing research, we plan to expand on the mean-reverting model for payout rates and (i) derive the optimal purchasing strategy assuming the targeted funds are earning interest or stochastic returns, (ii) derive the optimal allocation to DIAs within a more general portfolio context, and (iii) refine the asymptotic approximation to broaden the range of parameters within which our iterative scheme can be applied with minimal error. Another item on our to-do list is to provide some further theoretical justification for the workhorse we built this article on, namely Equation (3) and perhaps calibrate and estimate the parameters themselves, once a suitable dataset becomes available. Indeed, this is all part of ambitious plan and we suspect market annuity payout rates will have exited the continuation region by the time our research agenda is completed.

Appendices

Appendix A: Proofs of Propositions 4 and 5

Appendix A.1: Proof of Proposition 4

Let \hat{f} and $\hat{\pi}$ denote the σ^2 -order approximations of f and π^* , respectively, and formally write

$$\hat{f}(\pi, t) = f^{(0)}(\pi, t) + \sigma^2 f^{(1)}(\pi, t),$$

and

$$\hat{\pi}(t) = \pi^{(0)}(t) + \sigma^2 \pi^{(1)}(t),$$

in which $f^{(0)}$, $f^{(1)}$, $\pi^{(0)}$, and $\pi^{(1)}$ are independent of σ .

From the variational inequality in Equation (16) and from the relationship between variational inequalities and free-boundary problems (see, e.g., Peskir and Shiryaev, 2006), on the closure of the continuation region, f solves the following free-boundary problem:

$$\begin{cases} f_t + \kappa(\bar{\pi}(t) - \pi) f_\pi + \frac{1}{2} \sigma^2 \pi^2 f_{\pi\pi} = 0, \\ f(\pi^*(t), t) = \pi^*(t), \quad f_\pi(\pi^*(t), t) = 1, \\ f(\pi, T) = \pi, \quad 0 \leq \pi < \pi^*(T-). \end{cases} \quad (\text{A.1})$$

The smooth-fit condition $f_\pi(\pi^*(t), t) = 1$ comes from the optimality of the free boundary (Peskir and Shiryaev, 2006). Also, because π^* is differentiable with respect to t on $[0, T)$ (Friedman, 1975), if we differentiate $f(\pi^*(t), t) = \pi^*(t)$ fully with respect to t , we obtain the condition that $f_t(\pi^*(t), t) = 0$ on $[0, T)$.

Substitute \hat{f} and $\hat{\pi}$ for f and π^* , respectively, in the free-boundary problem in Equation (A.1), including the condition $f_t(\pi^*(t), t) = 0$; then, collect terms of order σ^0 versus σ^2 to obtain two free-boundary problems. To demonstrate a part of this process, first expand the differential equation for f up to order σ^2 .

$$\left(f_t^{(0)} + \sigma^2 f_t^{(1)} \right) + \kappa(\bar{\pi}(t) - \pi) \left(f_\pi^{(0)} + \sigma^2 f_\pi^{(1)} \right) + \frac{1}{2} \sigma^2 \pi^2 f_{\pi\pi}^{(0)} = 0.$$

By collecting terms of like order in this expansion, we obtain

$$f_t^{(0)} + \kappa(\bar{\pi}(t) - \pi) f_\pi^{(0)} = 0,$$

and

$$f_t^{(1)} + \kappa(\bar{\pi}(t) - \pi) f_\pi^{(1)} + \frac{1}{2} \pi^2 f_{\pi\pi}^{(0)} = 0.$$

Next, expand the free-boundary condition $f(\pi^*(t), t) = \pi^*(t)$ up to order σ^2 .

$$\left(f^{(0)}(\pi^{(0)}, t) + \sigma^2 \pi^{(1)} f_\pi^{(0)}(\pi^{(0)}, t) \right) + \sigma^2 f^{(1)}(\pi^{(0)}, t) = \pi^{(0)} + \sigma^2 \pi^{(1)}.$$

Again, by collecting terms of like order, we obtain

$$f^{(0)}(\pi^{(0)}, t) = \pi^{(0)},$$

and

$$\pi^{(1)} f_\pi^{(0)}(\pi^{(0)}, t) + f^{(1)}(\pi^{(0)}, t) = \pi^{(1)}.$$

By continuing this procedure for the remaining free-boundary and terminal conditions, and by simplifying the boundary conditions where possible, we obtain the following two free-boundary problems. For ease of reference, we include (redundant) free-boundary conditions that we obtained by differentiating the value-matching condition in Equation (A.1).

$$\begin{cases} f_t^{(0)} + \kappa(\bar{\pi}(t) - \pi) f_\pi^{(0)} = 0, \\ f^{(0)}(\pi^{(0)}, t) = \pi^{(0)}, \quad f_\pi^{(0)}(\pi^{(0)}, t) = 1, \quad f_t^{(0)}(\pi^{(0)}, t) = 0, \\ f^{(0)}(\pi, T) = \pi, \quad 0 \leq \pi < \pi^{(0)}(T-), \end{cases} \tag{A.2}$$

and

$$\begin{cases} f_t^{(1)} + \kappa(\bar{\pi}(t) - \pi) f_\pi^{(1)} + \frac{1}{2} \pi^2 f_{\pi\pi}^{(0)} = 0, \\ f^{(1)}(\pi^{(0)}, t) = 0, \quad \pi^{(1)} f_{\pi\pi}^{(0)}(\pi^{(0)}, t) + f_\pi^{(1)}(\pi^{(0)}, t) = 0, \\ \pi^{(1)} f_{t\pi}^{(0)}(\pi^{(0)}, t) + f_t^{(1)}(\pi^{(0)}, t) = 0, \\ f^{(1)}(\pi, T) = 0. \end{cases} \tag{A.3}$$

In Equations (A.2) and (A.3), and henceforth, $\pi^{(0)} = \pi^{(0)}(t)$ and $\pi^{(1)} = \pi^{(1)}(t)$.

We begin by solving the problem in Equation (A.2). By substituting $\pi = \pi^{(0)}(t)$ in the differential equation and by using the conditions $f_\pi^{(0)}(\pi^{(0)}, t) = 1$ and $f_t^{(0)}(\pi^{(0)}, t) = 0$, we learn that

$$\pi^{(0)}(t) = \bar{\pi}(t), \quad 0 \leq t < T. \tag{A.4}$$

The differential equation is of first order in t and π , so we use the method of characteristics to find its general solution. To that end, first solve for the characteristic:

$$\frac{d\pi}{dt} = \kappa(\bar{\pi}(t) - \pi),$$

which implies that

$$\pi(t) = \pi(t)e^{\kappa t} - \kappa \int_0^t \bar{\pi}(s)e^{\kappa s} ds. \tag{A.5}$$

Thus, if we define

$$\xi = \pi e^{\kappa t} - \kappa \int_0^t \bar{\pi}(s)e^{\kappa s} ds, \tag{A.6}$$

and

$$\tilde{f}(\xi, t) = f^{(0)}(\pi, t), \tag{A.7}$$

then the differential equation becomes $\tilde{f}_t = 0$, whose solution equals $\tilde{f}(\xi, t) = F(\xi)$ for an arbitrary function F of ξ . We determine F via the conditions $f^{(0)}(\bar{\pi}(t), t) = \bar{\pi}(t)$ for $0 \leq t < T$, and $f^{(0)}(\pi, t) = \pi$ for $0 \leq \pi < \bar{\pi}(T)$. Thus, we have the following expression for $f^{(0)}$; see Remark A.1 below for a discussion of $f^{(0)}$.

$$f^{(0)}(\pi, t) = \begin{cases} \pi e^{-\kappa(T-t)} + \kappa \int_t^T \bar{\pi}(s)e^{-\kappa(T-s)} ds, & \text{if } 0 \leq \pi < \pi_\ell(t), \\ \bar{\pi}(t'), & \text{if } \pi_\ell(t) \leq \pi \leq \bar{\pi}(t), \end{cases} \tag{A.8}$$

in which π_ℓ is given by

$$\pi_\ell(t) = \bar{\pi}(T)e^{\kappa(T-t)} - \kappa \int_t^T \bar{\pi}(s)e^{\kappa(s-t)} ds, \tag{A.9}$$

and in which t' uniquely solves

$$\bar{\pi}(t')e^{\kappa t'} - \kappa \int_0^{t'} \bar{\pi}(s)e^{\kappa s} ds = \pi e^{\kappa t} - \kappa \int_0^t \bar{\pi}(s)e^{\kappa s} ds. \tag{A.10}$$

Note that the left side of Equation (A.10) is a decreasing function of t' ; thus, we can solve Equation (A.10) for $t' \in [t, T]$ if $\pi_\ell(t) \leq \pi \leq \bar{\pi}(t)$. If $\pi = \pi_\ell(t)$, then $t' = T$; if $\pi = \bar{\pi}(t)$, then $t' = t$.

Remark A.1. If the payout yield π at time t is close enough to $\bar{\pi}(t)$, as measured by $\pi_\ell(t) \leq \pi \leq \bar{\pi}(t)$, then t' is the time at which the payout yield hits $\bar{\pi}$ if $\sigma = 0$. In this case, the value of $f^{(0)}$ equals $\bar{\pi}(t')$, as seen in the second expression in Equation (A.8). On the other hand, if the payout yield π at time t is far away from $\bar{\pi}$, as measured by $0 \leq \pi < \pi_\ell(t)$, then the payout yield will not reach $\bar{\pi}$ before time T , and the individual will annuitize all his/her wealth at time T . In this case, the value of $f^{(0)}$ equals the value of the payout yield at time T , which is the first expression in Equation (A.8).

To determine $\pi^{(1)}(t)$, in the approximation $\hat{\pi}(t)$ of $\pi^*(t)$, we need $f_{\pi\pi}^{(0)}(\bar{\pi}(t), t)$, which we compute from $f^{(0)}(\pi, t)$ in Equation (A.8) when $\pi_\ell(t) \leq \pi \leq \bar{\pi}(t)$. From Equation (A.10), we obtain

$$\frac{\partial}{\partial \pi} \bar{\pi}(t') = e^{-\kappa(t'-t)},$$

and

$$\frac{\partial^2}{\partial \pi^2} \bar{\pi}(t') = -e^{-2\kappa(t'-t)} \frac{\kappa}{\bar{\pi}'(t')};$$

thus,

$$f_{\pi\pi}^{(0)}(\bar{\pi}(t), t) = \frac{\partial^2}{\partial \pi^2} \bar{\pi}(t')|_{t'=t} = -\frac{\kappa}{\bar{\pi}'(t)} = \frac{\kappa}{(\bar{\gamma} + \lambda_x(t))\bar{\pi}(t)}. \tag{A.11}$$

Differentiate $f^{(0)}$'s differential equation with respect to π to get

$$f_{t\pi}^{(0)} + \kappa(\bar{\pi}(t) - \pi)f_{\pi\pi}^{(0)} - \kappa f_\pi^{(0)} = 0.$$

Set $\pi = \bar{\pi}(t)$ in this equation and use $\pi^{(1)}f_{t\pi}^{(0)}(\bar{\pi}(t), t) + f_t^{(1)}(\bar{\pi}(t), t) = 0$ from Equation (A.3) and $f_\pi^{(0)}(\bar{\pi}(t), t) = 1$ from Equation (A.2). Recall that $\pi^{(0)}(t) = \bar{\pi}(t)$.

$$f_t^{(1)}(\bar{\pi}(t), t) = -\kappa \pi^{(1)}(t). \tag{A.12}$$

In $f^{(1)}$'s differential equation, set $\pi = \bar{\pi}(t)$ and substitute for $f_{\pi\pi}^{(0)}(\bar{\pi}(t), t)$ from Equation (A.11) and for $f_t^{(1)}(\bar{\pi}(t), t)$ from Equation (A.12) to obtain

$$\pi^{(1)}(t) = \frac{1}{2} \frac{\bar{\pi}(t)}{\bar{\gamma} + \lambda_x(t)}.$$

Thus, we have proved the σ^2 -order approximation of the free boundary $\pi^*(t)$ given in Equation (21).

Appendix A.2: Proof of Proposition 5

For a given value $\alpha > 0$, define the differential operator L^α on an appropriate set of functions by $L^\alpha g = Lg - \alpha g$, in which Lg is given in Equation (7). Consider the function $f^{\alpha,k}$ given by

$$f^{\alpha,k}(\pi, t) = e^{-\alpha(T-t)} \bar{f}(\pi, t) + k\sigma^2,$$

in which \bar{f} solves Equation (23) and k is a positive constant.

If $0 \leq \pi < \min(\hat{\pi}(t), \pi^*(t))$, then

$$L^\alpha \left(f^{\alpha,k} - e^{-\alpha(T-t)} f \right) = -\alpha k \sigma^2 < 0.$$

If $\hat{\pi}(t) \leq \pi < \pi^*(t)$, then

$$L^\alpha \left(f^{\alpha,k} - e^{-\alpha(T-t)} f \right) = -\alpha k \sigma^2 + e^{-\alpha(T-t)} \kappa(\bar{\pi}(t) - \pi) < 0,$$

in which we use the fact that $\hat{\pi}(t) > \bar{\pi}(t)$. If $\pi^*(t) \leq \pi < \hat{\pi}(t)$, then

$$L^\alpha \left(f^{\alpha,k} - e^{-\alpha(T-t)} f \right) = -\alpha k \sigma^2 - e^{-\alpha(T-t)} \kappa(\bar{\pi}(t) - \pi),$$

which is negative if $k \geq \frac{\kappa}{\alpha} \max_{0 \leq t \leq T} \pi^{(1)}(t)$. If $\pi \geq \max(\hat{\pi}(t), \pi^*(t))$, then

$$f^{\alpha,k} - e^{-\alpha(T-t)} f = \alpha k \sigma^2 > 0.$$

To summarize, we have $f^{\alpha,k} - e^{-\alpha(T-t)} f > 0$ if $\pi \geq \max(\hat{\pi}(t), \pi^*(t))$, $L^\alpha \left(f^{\alpha,k} - e^{-\alpha(T-t)} f \right) < 0$, if $k \geq \frac{\kappa}{\alpha} \max_{0 \leq t \leq T} \pi^{(1)}(t)$ and if $\pi < \max(\hat{\pi}(t), \pi^*(t))$, and $f^{\alpha,k}(\pi, T) - f(\pi, T) > 0$. It follows from the maximum–minimum principle (Walter, 1970, section 26) that

$$f^{\alpha,k} > e^{-\alpha(T-t)} f,$$

on $\mathbf{R}^+ \times [0, T]$ for all $\alpha > 0$ and $k \geq \frac{\kappa}{\alpha} \max_{0 \leq t \leq T} \pi^{(1)}(t)$, or equivalently,

$$f < \bar{f} + \frac{\kappa e^{\alpha(T-t)}}{\alpha} \max_{0 \leq t \leq T} \pi^{(1)}(t) \sigma^2.$$

As a function of α , the expression $\frac{e^{\alpha(T-t)}}{\alpha}$ is minimum when $\alpha = \frac{1}{T-t}$. Thus, we have

$$f < \bar{f} + T\kappa \max_{0 \leq t \leq T} \pi^{(1)}(t) \sigma^2,$$

and we have proved Proposition 5 with $\beta = T\kappa \max_{0 \leq t \leq T} \pi^{(1)}(t)$.

Appendix B: Proof of Proposition 7

To make analysis of the free boundary more tractable, define G by

$$G(z, \pi, t) = \ln g(z, \pi, t).$$

Then, G solves the following free-boundary problem.

$$\begin{cases} G_t + \kappa(\bar{\pi}(t) - \pi) G_\pi + \frac{1}{2} \sigma^2 \pi^2 (G_\pi^2 + G_{\pi\pi}) = 0, \\ G_z(z, \pi^*(z, t), t) = (1 - \gamma) \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}, \\ G_{zz}(z, \pi^*(z, t), t) = -(1 - \gamma) \left(\frac{\pi^*(z, t)}{1 + \pi^*(z, t)z} \right)^2, \\ G(0, \pi, t) = 0, \\ G(z, \pi, T) = (1 - \gamma) \ln(1 + \pi z), \quad 0 \leq \pi < \pi^*(z, T-). \end{cases} \tag{A.13}$$

Furthermore, because the free-boundary conditions are expressed in terms of G_z , we use G_z 's free-boundary problem to obtain an approximation of $\pi^*(z, t)$.

$$\begin{cases} G_{zt} + \kappa(\bar{\pi}(t) - \pi) G_{z\pi} + \frac{1}{2} \sigma^2 \pi^2 (2G_{\pi}G_{z\pi} + G_{z\pi\pi}) = 0, \\ G_z(z, \pi^*(z, t), t) = (1 - \gamma) \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}, \\ G_{zz}(z, \pi^*(z, t), t) = -(1 - \gamma) \left(\frac{\pi^*(z, t)}{1 + \pi^*(z, t)z} \right)^2, \\ G_z(z, \pi, T) = (1 - \gamma) \frac{\pi}{1 + \pi z}, \quad 0 \leq \pi < \pi^*(z, T-). \end{cases} \tag{A.14}$$

Fully differentiate $G_z(z, \pi^*(z, t), t) = (1 - \gamma) \frac{\pi^*(z, t)}{1 + \pi^*(z, t)z}$ with respect to z to obtain

$$G_{z\pi}(z, \pi^*(z, t), t) = \frac{1 - \gamma}{(1 + \pi^*(z, t)z)^2}. \tag{A.15}$$

Then, fully differentiate the same free-boundary condition with respect to t and simplify the result to obtain

$$G_{zt}(z, \pi^*(z, t), t) = 0. \tag{A.16}$$

In this appendix, we approximate the optimal exercise boundary $\pi^*(z, t)$. The approximation that we obtain also holds when $\gamma = 1$ (i.e., logarithmic utility), but obtaining it requires a different argument, which we do not include for the sake of space. Also, as $\gamma \rightarrow 0+$, the approximation approaches the one in Equation (21) for the risk-neutral purchaser.

Let \hat{G} and $\hat{\pi}$ denote the σ^2 -order approximations of G and π^* , respectively, and formally write

$$\hat{G}(z, \pi, t) = G^{(0)}(z, \pi, t) + \sigma^2 G^{(1)}(z, \pi, t),$$

and

$$\hat{\pi}(z, t) = \pi^{(0)}(z, t) + \sigma^2 \pi^{(1)}(z, t),$$

in which $G^{(0)}$, $G^{(1)}$, $\pi^{(0)}$, and $\pi^{(1)}$ are independent of σ . For simplicity, in what follows, we write $\pi^{(0)}$ and $\pi^{(1)}$ without the arguments z and t . As in Appendix A.1, substitute these approximations for G and for π^* into Equation (A.14) to get the following two free-boundary problems. For ease of reference, we include (redundant) free-boundary conditions that we obtained by differentiating the smooth-fit condition in Equation (A.14).

$$\begin{cases} G_{zt}^{(0)} + \kappa(\bar{\pi}(t) - \pi) G_{z\pi}^{(0)} = 0, \\ G_z^{(0)}(z, \pi^{(0)}, t) = (1 - \gamma) \frac{\pi^{(0)}}{1 + \pi^{(0)}z}, \\ G_{zz}^{(0)}(z, \pi^{(0)}, t) = -(1 - \gamma) \left(\frac{\pi^{(0)}}{1 + \pi^{(0)}z} \right)^2, \\ G_{z\pi}^{(0)}(z, \pi^{(0)}, t) = \frac{1 - \gamma}{(1 + \pi^{(0)}z)^2}, \quad G_{zt}^{(0)}(z, \pi^{(0)}, t) = 0, \\ G_z^{(0)}(z, \pi, T) = (1 - \gamma) \frac{\pi}{1 + \pi z}, \quad 0 \leq \pi < \pi^{(0)}(z, T-). \end{cases} \tag{A.17}$$

and

$$\left\{ \begin{array}{l} G_{zt}^{(1)} + \kappa(\bar{\pi}(t) - \pi) G_{z\pi}^{(1)} + \frac{1}{2} \pi^2 (2G_{\pi}^{(0)} G_{z\pi}^{(0)} + G_{z\pi\pi}^{(0)}) = 0, \\ G_z^{(1)}(z, \pi^{(0)}, t) = 0, \\ \pi^{(1)} G_{z\pi}^{(0)}(z, \pi^{(0)}, t) + G_{z\pi}^{(1)}(z, \pi^{(0)}, t) = -2(1 - \gamma) \frac{\pi^{(1)} \pi^{(0)}}{(1 + \pi^{(0)} z)^3}, \\ \pi^{(1)} G_{z\pi\pi}^{(0)}(z, \pi^{(0)}, t) + G_{z\pi\pi}^{(1)}(z, \pi^{(0)}, t) = -2(1 - \gamma) \frac{\pi^{(1)} z}{(1 + \pi^{(0)} z)^3}, \\ \pi^{(1)} G_{zt\pi}^{(0)}(z, \pi^{(0)}, t) + G_{zt\pi}^{(1)}(z, \pi^{(0)}, t) = 0, \\ G_z^{(1)}(z, \pi, T) = 0, \quad 0 \leq \pi < \pi^*(z, T-). \end{array} \right. \quad (\text{A.18})$$

If we substitute $\pi = \pi^{(0)}(z, t)$ into $G_z^{(0)}$'s differential equation and use the condition $G_{zt}^{(0)}(z, \pi^{(0)}, t) = 0$, then we get

$$\pi^{(0)}(z, t) = \bar{\pi}(t), \quad (\text{A.19})$$

as in the risk-neutral case; see Equation (A.4). Then, by using the method of characteristics as in Appendix A.1, we deduce that for $\pi_\ell(t) \leq \pi \leq \bar{\pi}(t)$,

$$G_z^{(0)}(z, \pi, t) = (1 - \gamma) \frac{\bar{\pi}(t')}{1 + \bar{\pi}(t')z}, \quad (\text{A.20})$$

in which $\pi_\ell(t)$ and t' are given in Equations (A.9) and (A.10), respectively. Next, differentiate $G_z^{(0)}$'s differential equation with respect to π .

$$G_{z\pi}^{(0)} + \kappa(\bar{\pi}(t) - \pi) G_{z\pi\pi}^{(0)} - \kappa G_{z\pi}^{(0)} = 0.$$

Set $\pi = \bar{\pi}(t)$ in this equation and use $\pi^{(1)} G_{z\pi}^{(0)}(z, \bar{\pi}(t), t) + G_{z\pi}^{(1)}(z, \bar{\pi}(t), t) = 0$ from Equation (A.18) and $G_{z\pi}^{(0)}(z, \bar{\pi}(t), t) = \frac{1-\gamma}{(1+\bar{\pi}(t)z)}$ from (A.17).

$$G_{z\pi}^{(1)}(z, \bar{\pi}(t), t) = -\kappa(1 - \gamma) \frac{\pi^{(1)}}{(1 + \bar{\pi}(t)z)^2}. \quad (\text{A.21})$$

In $G_z^{(1)}$'s differential equation, set $\pi = \bar{\pi}(t)$, as per Appendix B, and use $G_{z\pi}^{(0)}(z, \bar{\pi}(t), t)$ from (B.5) to get

$$\kappa(1 - \gamma) \frac{\pi^{(1)}}{(1 + \bar{\pi}(t)z)^2} = \frac{1}{2} \bar{\pi}^2(t) \left(\frac{2(1 - \gamma)}{(1 + \bar{\pi}(t)z)^2} G_{\pi}^{(0)}(z, \bar{\pi}(t), t) + G_{z\pi\pi}^{(0)}(z, \bar{\pi}(t), t) \right). \quad (\text{A.22})$$

From $G_z^{(0)}$ in (A.20) and from (A.10), we get

$$G_{z\pi\pi}^{(0)}(z, \bar{\pi}(t), t) = -(1 - \gamma) \left(\frac{2z(1 - \gamma)}{(1 + \bar{\pi}(t)z)^3} + \frac{\kappa}{(1 + \bar{\pi}(t)z)^2} \frac{1}{\bar{\pi}'(t)} \right). \quad (\text{A.23})$$

Getting $G_{\pi}^{(0)}(z, \bar{\pi}(t), t)$ is a bit more involved. Because $\bar{\pi}(t')$ is independent of z , we can integrate $G_z^{(0)}$ to obtain

$$G^{(0)}(z, \pi, t) = (1 - \gamma) \ln(1 + \bar{\pi}(t')z),$$

in which we use $G^{(0)}(0, \pi, t) = 0$, which comes from the condition $G(0, \pi, t) = 0$ in (A.13). Differentiate this expression of $G^{(0)}$ with respect to π .

$$G_{\pi}^{(0)}(z, \pi, t) = (1 - \gamma) \frac{z}{1 + \bar{\pi}(t)z} e^{-\kappa(t'-t)},$$

which implies that

$$G_{\pi}^{(0)}(z, \bar{\pi}(t), t) = (1 - \gamma) \frac{z}{1 + \bar{\pi}(t)z}. \tag{A.24}$$

If we substitute the expression of $G_{z\pi\pi}^{(0)}(z, \bar{\pi}(t), t)$ from (A.23) and $G_{\pi}^{(0)}(z, \bar{\pi}(t), t)$ from (A.24) into (A.22), we get the following expression for $\pi^{(1)}$.

$$\pi^{(1)}(z, t) = \bar{\pi}(t) \left(\frac{1}{2} \frac{1}{\bar{r} + \lambda_x(t)} - \frac{\gamma}{\kappa} \frac{\bar{\pi}(t)z}{1 + \bar{\pi}(t)z} \right).$$

Thus, we have proved the σ^2 -order approximation of the free boundary $\pi^*(z, t)$ in Proposition 7.

Appendix C: Verification of the σ^2 -Order Approximate Solutions

To verify the σ^2 -order approximate solutions in Propositions 4 and 7, we compare them to the numerical solution of the variational inequality (11) with terminal condition $g(z, \pi, T) = (1 + \pi z)^{1-\gamma}$ and boundary condition $g(0, \pi, t) = 1$.

We discretize the variational inequality (11) and obtain

$$\max \left((1 - \gamma) L_b g_{j,k}^i, \pi_j g_{j,k-1}^i - \frac{1 + \pi_j z_{k-1}}{1 - \gamma} D_{z,b} g_{j,k}^i \right) = 0, \tag{A.25}$$

with $g_{j,k}^M = (1 + \pi_j z_k)^{1-\gamma}$ and $g_{j,1}^i = 1$. Here $g_{j,k}^i$ is the approximate to $g(z, \pi, t)$ on the grid (z_k, π_j, t_i) , and the discrete operators are given as

$$L_b g_{j,k}^i = \frac{g_{j,k}^i - g_{j,k}^{i-1}}{\delta t} + c_{j,k}^i D_{\pi,b} g_{j,k}^i + \frac{\sigma^2 \pi_j^2}{2} D_{\pi,b}^2 g_{j,k}^i,$$

in which $c_{j,k}^i = \kappa(\bar{\pi}_i - \pi_j)$,

$$D_{\pi,b} g_{j,k}^i = \frac{c_{j,k}^i + |c_{j,k}^i| g_{j,k}^i - g_{j-1,k}^i}{2 \delta \pi} + \frac{c_{j,k}^i - |c_{j,k}^i| g_{j+1,k}^i - g_{j,k}^i}{2 \delta \pi} \tag{A.26}$$

is the upwind scheme for the first-order derivative g_{π} ,

$$D_{\pi,b}^2 = \frac{g_{j+1,k}^i + g_{j-1,k}^i - 2g_{j,k}^i}{(\delta \pi)^2} \tag{A.27}$$

is the centre difference scheme for the derivative $g_{\pi\pi}$, and

$$D_{z,b} g_{j,k}^i = \frac{g_{j,k}^i - g_{j,k-1}^i}{\delta z}. \tag{A.28}$$

The grid values are given by $t_i = (i - 1)\delta t$, $\pi_j = (j - 1)\delta \pi$, $z_k = (k - 1)\delta z$ for $i = 1, \dots, M$, $j = 1, \dots, J$, and $k = 1, \dots, K$. The grid sizes are given by $\delta t = T/(M - 1)$, $\delta \pi = \Pi/(J - 1)$, and $\delta z = Z/(K - 1)$ and the computational domain is given by $[0, T] \times [0, \Pi] \times [0, Z]$.

We solve the discrete variational inequality backward in time, that is, $i = M, \dots, 1$, and forward in the variable z , that is, $k = 1, \dots, K$. As shown in Figure 5, the agreement between

the numerical (symbols) and σ^2 -order approximate (lines) solutions is excellent when $\sigma = 5\%$, $\kappa = 10\%$, and $\gamma = 5$. The computational domain is $[0, 20] \times [0, 2.5] \times [0, 8]$. Our results also show that σ^2 -order approximate solution becomes less accurate for larger values of σ and γ .

The numerical method for the risk-neutral case is simpler and is obtained using a similar method (without the z component) and it agrees with the solution for the risk-neutral case when $\gamma = 0$, as noted in Remark 12.

References

- Bayraktar, E. and Young, V. A. (2009) Minimizing the probability of lifetime ruin with deferred life annuities, *North American Actuarial Journal* **13**, 141–154.
- Benartzi, S., Previtiero, A., and Thaler, R. H. (2011) Annuitization puzzles, *Journal of Economic Perspectives* **25**, 143–164.
- Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., and Nesbitt, C. J. (1997) *Actuarial Mathematics*, 2nd edition, Society of Actuaries, Schaumburg, Illinois.
- Brennan, M. J. and Schwartz, E. S. (1977) The valuation of American put options, *Journal of Finance* **32**, 449–462.
- Cairns, A. J. G., Blake, D., and Dowd, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration, *Journal of Risk and Insurance* **73**, 687–718.
- Campbell, J. Y. and Viciera, L. M. (2002) *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press, Oxford.
- Charupat, N., Kamstra, M. J., and Milevsky, M. A. (2015) The sluggish and asymmetric reaction of annuity prices to changes in interest rates, *Journal of Risk and Insurance*, in press, doi:10.1111/jori.12061.
- Cocco, J. F., Franciso, J. G., and Pascal, J. M. (2005) Portfolio choice over the life-cycle, *Review of Financial Studies* **18**, 491–533.
- Davis, M. H. A. and Norman, A. R. (1990) Portfolio selection with transaction costs, *Mathematics of Operations Research* **15**, 676–713.
- Dixit, A. K. (1991) A simplified treatment of the optimal regulation of Brownian motion, *Journal of Economic Dynamics and Control* **15**, 657–673.
- Dixit, A. K. and Pindyck, R. S. (1994) *Investment under Uncertainty*, Princeton University Press, Princeton.
- Dumas, B. (1991) Super contact and related optimality conditions, *Journal of Economic Dynamics and Control* **15**, 675–685.
- Dushi, I. and Webb, A. (2004) Household annuitization decisions: simulations and empirical analyses, *Journal of Pension Economics and Finance* **3**, 109–143.
- Farhi, E. and Panageas, S. (2007) Saving and investing for early retirement: a theoretical analysis, *Journal of Financial Economics* **83**, 87–121.
- Friedman, A. (1975) Parabolic variational inequalities in one space dimension and smoothness of the free boundary, *Journal of Functional Analysis* **18**, 151–176.
- Hainuat, D. and Deelstra, G. (2014) Optimal timing for annuitization, based on jump-diffusion fund and stochastic mortality, *Journal of Economic Dynamics and Control* **44**, 124–146.
- Horneff, W. J., Maurer, R. H., Mitchell, O. S., and Stamos, M. Z. (2009) Asset allocation and location over the life cycle with investment-linked survival-contingent payouts, *Journal of Banking and Finance* **33**, 1688–1699.
- Inkman, J., Lopes, P., and Michaelides, A. (2011) How deep is the annuity market participation puzzle?, *Review of Financial Studies* **24**, 279–319.
- Kingston, G. and Thorp, S. (2005) Annuitization and asset allocation with HARA utility, *Journal of Pension Economics and Finance* **4**, 225–248.

- Koijen, R. S. J., Nijman, T. E., and Werker, B. J. M. (2011) Optimal annuity risk management, *Review of Finance* 15, 799–833.
- Koijen, R. S. J. and Yogo, M. (2015) The cost of financial frictions for life insurers, *American Economic Review* 105, 445–475.
- Konicz, A. K. and Mulvey, J. M. (2013) Applying a stochastic financial planning system to an individual: Immediate or deferred life annuities?, *The Journal of Retirement* 1, 46–60.
- Lee, R. D. and Carter, L. R. (1992) Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association* 87, 659–671.
- Milevsky, M. A., Huang, H., and Young, V. R. (2015) A glide path for target date annuitization, *Journal of Retirement* 3, 27–37.
- Milevsky, M. A. and Young, V. R. (2007) Annuitization and asset allocation, *Journal of Economic Dynamics and Control* 31, 3138–3177.
- Neuberger, A. (2003) Annuities and the optimal investment decision. Working paper, London Business School.
- Peskir, G. and Shiryaev, A. (2006) *Optimal Stopping and Free-Boundary Problems*, Birkhäuser Verlag, Basel.
- Previtero, A. (2014) Stock market returns and annuitization, *Journal of Financial Economics* 113, 202–214.
- Richard, S. F. (1975) Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model, *The Journal of Financial Economics* 2, 187–203.
- Scott, J. S., Watson J. G., and Hu, W.-Y. (2011) What makes a better annuity?, *Journal of Risk and Insurance* 78, 213–244.
- Stabile, G. (2006) Optimal timing of the annuity purchase: combined stochastic control and optimal stopping problem, *International Journal of Theoretical and Applied Finance* 9, 151–170.
- Walter, W. (1970) *Differential and Integral Inequalities*, Springer-Verlag, New York.
- Wang, T. and Young, V. R. (2012) Maximizing the utility of consumption with commutable annuities, *Insurance: Mathematics and Economics* 51, 352–369.
- Yaari, M. (1965) Uncertain lifetime, life insurance and the theory of the consumer, *Review of Economic Studies* 32, 137–150.