

Annuitization and asset allocation with HARA utility

GEOFFREY KINGSTON

School of Economics, University of New South Wales, Sydney
(e-mail: g.kingston@unsw.edu.au)

SUSAN THORP

School of Finance and Economics, University of Technology, Sydney
(e-mail: susan.thorp@uts.edu.au)

Abstract

A new explanation for the well-known reluctance of retirees to buy life annuities is due to Milevsky and Young (2002, 2003): Since the decision to purchase longevity insurance is largely irreversible, in uncertain environments a real option to delay annuitization (RODA) generally has value. Milevsky and Young analytically identify and numerically estimate the RODA in a setting of constant relative risk aversion. This paper presents an extension to the case of HARA (or GLUM) preferences, the simplest representation of a consumption habit. The precise date of annuitization can no longer be ascertained with certainty in advance. This paper derives an approximation whereby the agent precommits. The effect of increasing the subsistence consumption rate on the timing of annuity purchase is similar to the effect of increasing the curvature parameter of the utility function. As in the CRRA case studied by Milevsky and Young, delayed annuitization is associated with optimistic predictions of the Sharpe ratio and divergence between annuity purchaser and provider predictions of mortality.

1 Introduction

One of the most important financial decisions many people make is the choice of a portfolio of assets during retirement. One difficult decision concerns longevity risk, where individuals face the possibility of outliving their resources or, alternatively, of foregoing consumption by dying before wealth is exhausted. Economic theory has long maintained that the protection against longevity risk offered by annuities is valuable and should therefore be a sought-after product (Yaari, 1965; Davidoff, Brown, and Diamond, 2003). But, despite the ready availability of longevity insurance, retirees across the world seldom voluntarily annuitize.

We thank Hazel Bateman, Henry Jin, Moshe Milevsky, Sachi Purcal, Mike Sherris, Emil Valdez, Virginia Young, anonymous referees and conference participants at the 3rd IFID Centre Conference on Asset Allocation and Mortality, University of Toronto, the 11th Australian Colloquium of Superannuation Researchers, University of New South Wales, and the Australasian meeting of the Econometric Society 2004, for comments and advice. Plan for Life responded generously to our requests for data. Financial assistance from the Australian Research Council is gratefully acknowledged.

Milevsky and Young (2002, 2003) have come up with a new explanation for the reluctance of retirees to buy life annuities. They begin with the observation that negotiable annuity contracts are not available in general, so that the decision to purchase longevity insurance is largely irreversible. The literature on real options demonstrates that in an uncertain environment it often pays to delay investments that cannot easily be reversed. Using a Merton (1969) continuous-time model under constant relative risk aversion (CRRA), together with Ibbotson Associates financial data and North American mortality data, Milevsky and Young find that it is generally better to delay buying a fixed life annuity until age 70. The intuition is straightforward: a longer period of exposure to the risky asset prior to annuitization offers people a chance to improve their budget constraint that evaporates after annuitization. So even risk averse individuals may decide to delay in the expectation of creating more wealth and enjoying a higher long-term income. Further, it has also been generally thought that people will avoid buying annuities whenever their personal evaluation of life expectancy is lower than average. But Milevsky and Young demonstrate that this perception is too limited. They show that individuals who expect to live longer than average may also delay annuitizing, anticipating that they will survive to benefit from risky asset exposure and falling annuity premiums as they age.

Our aim is to extend and reassess their analysis for a more general description of preferences. Hyperbolic absolute risk aversion, otherwise known as linear risk tolerance, was introduced to continuous-time modeling by Merton (1971). A one-parameter special case can be obtained by setting the curvature parameter of the utility function equal to one; that is, by defining one-period utility as the logarithm of consumption plus a constant. This special case has long been known to micro-economists as the Stone-Geary utility function, and has been described more recently by financial economists as the generalized logarithmic utility model (GLUM) (Rubinstein, 1976). The version of HARA preferences used in our set-up measures utility over consumption relative to a predetermined floor or subsistence level.

There are several reasons why the Milevsky-Young analysis is worth extending to the HARA (or GLUM) case. First, a fixed consumption floor is the simplest possible representation of a consumption habit, either external or internal. Insofar as the habit paradigm constitutes a useful characterization of any phase of the life-cycle, surely retirement is that phase. Second, the investor with CRRA preferences buys in falling markets and sells in rising ones. Perold and Sharpe (1988) point out, however, that only a buy-and-hold strategy is consistent with the behavior of the 'average' investor in equilibrium. Stone-Geary (or GLUM) preferences can capture buy-and-hold behavior.¹ Third, HARA preferences can capture a taste for 'portfolio insurance', whereby the investor seeks a convex payoff profile and implements this in part by buying in rising markets and selling in falling ones.² Finally, HARA (or GLUM) may

¹ For evidence that the investment behaviour of the elderly shows buy-and-hold behavior, see Ameriks and Zeldes (2001).

² Leland (1980) provides a two-period analysis of conditions on the value function that generate a demand for portfolio insurance. Kingston (1989) discusses conditions under which HARA utility generates a demand for portfolio insurance of the 'constant proportion' variety. Having closed off the option to continue working, retirees no longer hold embedded put options on stocks (Liu and Neiss, 2002) and therefore can be considered more likely than workers to be buyers of portfolio insurance.

be superior to CRRA when a financial planner is attempting to elicit information about the risk tolerance of a retired client. In particular, the planner may find it easier to phrase clarifying questions in terms of the client's minimum consumption requirements than the usual questions about indifference between choices involving hypothetical gambles.

Readers of Ingersoll's (1987) classic text will know that the key to solving problems involving HARA utility is to transform the state variable for wealth so as to reduce the problem to one of CRRA utility with a state variable net of an 'escrowed' wealth component that protects the consumption floor. Two escrow funds are needed in the present case, one to protect floor consumption prior to annuitization, and the other to protect floor consumption afterwards. We demonstrate that the presence of these two escrow funds reduces the optimal delay period, bringing forward annuitization whenever the consumption floor is non-zero. In addition, numerical examples show that moderately risk averse investors who wish to insure 50 per cent of their consumption stream will optimally annuitize 5-6 years earlier than those who do not. The analysis also shows that divergence between the annuitant's subjective judgement of life expectancy and the annuity provider's objective judgement will increase the optimal delay, as Milevsky and Young demonstrated, but the range of divergence is more compressed whenever the consumption floor is non-zero.

The evidence on voluntary annuitization is reviewed in Section 2, with particular reference to the Australian case. The theoretical derivation for the optimal timing of annuitization for an individual with HARA preferences is set out in Section 3. Section 4 presents numeric illustrations and Section 5 concludes. Appendices A to D contain algebraic derivations that are too long for the main text.

2 Voluntary annuitization patterns

Reluctance to buy life annuities is a worldwide phenomenon.³ The standard explanations include: high actuarial loadings arising from adverse selection, the wish to make bequests to heirs, alternative support from family members or from life income streams provided by the government, the wish to self-insure against the contingencies of expensive health care or nursing home care, high life-office margins arising from incompleteness in either the maturity structure or the contingency structure of government bonds on issue, and inadequate consumer education.⁴ Long as this list is, it has not proved wholly convincing and the empirical puzzle persists.

Low levels of voluntary annuitization are becoming more evident as countries add defined-contribution components to existing retirement savings schemes. In Australia, for example, where all employees over the age of 18 years contribute a mandatory percentage of their earnings to retirement savings, no-one is compelled to purchase longevity insurance at retirement. A means-tested government pension provides a real annuity to qualifying people over 65 years of age. For other retirees, an array of tax-preferred income stream choices are on offer. These include immediate fixed annuities (life and term), immediate variable term annuities ('term allocated

³ Milevsky and Young (2002) provide relevant statistics for the US case.

⁴ Mitchell and McCarthy (2002) provide a detailed account of these issues.

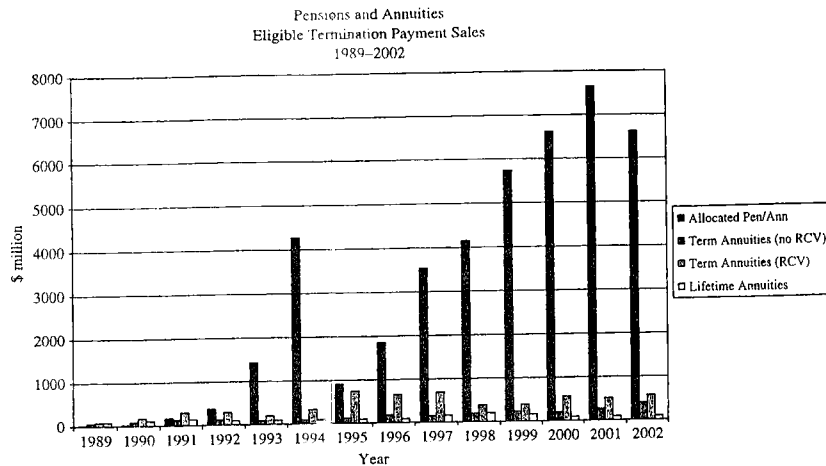


Figure 1. Australian sales of retirement income streams
Source: Plan For Life (2003).

pensions'), and phased withdrawal products ('allocated pensions'). Just as in other parts of the developed world, life annuities are not a popular choice in Australia. In 2002 Australians held assets of \$11.6 billion in the form of life annuities (IFSA, 2002). Compared to the \$34 billion invested in allocated pensions, this was a modest amount.⁵

Australian regulators have tried unsuccessfully to motivate annuity purchases.⁶ When an individual allocates at least half of their retirement savings to a life or life-expectancy annuity with no residual capital value, they are allowed a higher tax-concessional income, and the value of the annuity is also treated concessionally in social security means tests. Despite this, Figure 1 below shows that people have seldom purchased life annuities. Instead, most retirement savings are held in allocated pension accounts, which have regulated withdrawal limits, but offer no risk pooling. In addition, the average allocated pension account maintains a 60 per cent exposure to risky asset classes – stocks and commercial real estate – suggesting that the prospect of better returns is an important factor in retirees' choices. The prospect of higher returns, combined with ongoing control over one's own portfolio has evidently been appealing.

⁵ Moreover, few of these would have involved purchased products, as distinct from non-marketed entitlements from defined-benefit pension plans. (See Figure 1.)

⁶ Australian regulators have recently introduced market-linked income stream products called 'term allocated pensions'. Their aim is to offer a non-commutable, variable income stream with a term of life-expectancy, but without the restricted portfolio base of conventional fixed life or term annuities. There appears to be no longevity-risk pooling feature in these products.

Sales of allocated pensions dominate sales of all fixed annuity types since 1989. Even when one considers immediate annuities separately, sales of lifetime products (very small to begin with) have continued to decline in favour of term annuities, particularly those which return residual capital. The insurance implicit in social security payments may account for at least some of this disparity, along with the impact of loadings and adverse selection.

3 Options to delay annuitization

Predictions from life-cycle theories depend on assumptions about agents' preferences for consumption and risk. The CRRA model assumes that agents derive satisfaction from the absolute level of their consumption. Rubinstein (1976) and others advocated the alternative view that utility from consumption was better measured relative to some reference level. In other words, utility increases only as consumption rises above a floor or subsistence. More recently, the habit formation literature (see Constantinides (1990) and Campbell and Cochrane (1999) for example) has generalized the idea of relative utility by allowing the consumption floor to vary over time according to an internal or external habit.

A generalized utility (HARA) model also meshes more naturally with the pensions policy debate. Consider the way people plan for retirement. The most common metric for the adequacy of an accumulation is the long-term income stream it can generate. Pension calculators frame retirement provision in terms of 'required gross income in today's dollars'. This question aims to identify the minimum consumption stream a person can adequately subsist on, which is also the basic idea behind discussion of replacement rates. To describe such a preference for subsistence consumption one needs a non-zero consumption floor in the utility function. Hence we work with a HARA utility function.

3.1 After annuitization

Annuitization at time T is taken to mean that a retiree aged $x+T$ places all her assets in a real life annuity.⁷ Following Milevsky and Young [M–Y] we assume for simplicity that the real interest rate r is equal to the rate of time preference. Combined with the assumed absence of risky assets in the life annuity, a consequence is that consumption after annuitization is level. The subjective hazard rate after annuitization is denoted by λ_{x+T+t}^s ($t \geq 0$). The corresponding subjective annuity factor for an agent aged $x+T$ is \bar{a}_{x+T}^s . The objective hazard rate after annuitization is λ_{x+T+t}^o and the corresponding objective probability of survival is ${}_t p_{x+T}^o$. In the case of time-separable HARA preferences, discounted direct utility at time t after annuitization is

$$U(C, t; T) = \frac{(C_t - \hat{C})^{1-\gamma}}{1-\gamma} e^{-\left(r + \int_{x+T}^{x+T+t} \lambda_v^s dv\right)t} \quad (1)$$

⁷ It appears that the only deep market anywhere in the world for variable (with-profit) life annuities is within the TIAA-CREF pension plan in the United States (see Milevsky and Young).

so that the corresponding equation for indirect utility is

$$\frac{\left(\frac{W_T}{\bar{a}_{x+T}^s} - \hat{C}\right)^{1-\gamma}}{1-\gamma} e^{-\left(rT + \int_{x+T}^{x+T+t} \lambda_x^s dv\right)} \tag{2}$$

where the actuarial present value of a life annuity paying one dollar per year is

$$\bar{a}_x = \int_0^\infty e^{-rt} p_x dt$$

where

$${}_t p_x = e^{-\int_0^t \lambda_{x+v} dv}$$

and \bar{a}_{x+T}^s is the market annuity factor for an individual aged $x+T$.

Seen from time zero the individual's value function at any given future time T is therefore given by

$$V(W, 0, T) \equiv e^{-rT} p_x^s \bar{a}_{x+T}^s E \left[\frac{\left(\frac{W_T}{\bar{a}_{x+T}^s} - \hat{C}\right)^{1-\gamma}}{1-\gamma} \right] \tag{4}$$

Equation (4) implies a boundary condition $V(W, T, T) = \bar{a}_{x+T}^s \left(\frac{W_T}{\bar{a}_{x+T}^s} - \hat{C}\right)^{1-\gamma} / 1-\gamma$ for the retiree's pre-annuitization PDE.

3.2 The pre-annuitization problem

Prior to annuitization ($0 \leq t < T$), the retiree holds her wealth in a portfolio invested partly in shares with instantaneous expected return α and variance σ^2 , and partly in indexed bonds with known return r .

The risky asset process is described by a conventional geometric Brownian motion

$$dS(t) = \alpha S(t) dt + \sigma S(t) dz_t$$

where dz_t is a standard Wiener process. In the numerical analysis below values for α and σ are set by reference to long-term forecasts of the real equity premium and observed values of equity market volatility. The indexed bond return is assumed to follow

$$dB(t) = rB(t) dt.$$

At time zero the retiree's problem is to make contingent plans for the amount Π , invested in stocks, along with consumption C_t , and a date T on which the retiree's wealth is annuitized

$$\max_{C_t, \Pi, T} E \left[\int_0^T e^{-rt} ({}_t p_x^s)^s \frac{(C_t - \hat{C})^{1-\gamma}}{1-\gamma} dt + V(W, 0, T) \right], \tag{5}$$

subject to

$$dW_t = [rW_t + (\alpha - r)\Pi_t - C_t] dt + \sigma \Pi_t dz_t,$$

where E denotes the expectations operator, and dz_t is a Wiener increment.

3.3 Solution

Optimal stopping problems comparable with the one here are analyzed by Karatzas and Wang (2000). They find that it is difficult to characterize the solution to problems for which the optimal stopping time changes stochastically with the evolution of the state variable, i.e. wealth. In contrast to the problem solved by Milevsky and Young, the problem studied here has this complication. It remains under the duality approach recommended by Karatzas and Wang, as distinct from the HJB equation of dynamic programming utilized by Milevsky and Young (and by the present paper). For tractability we require that the retiree chooses an optimal annuitization time at time zero. Put another way, we disregard the 'American put' feature of our problem. Analysis of this feature is left to future research.

This paper characterizes a precommitment to T on the basis of information available at any prior time, t . This first-order condition has the desirable property of reducing to the first-order condition derived by Milevsky and Young in the case of constant relative risk aversion, in other words, a zero consumption floor.

Consider a retiree who makes a pre-commitment to annuitize at a future date, T . Once a pre-commitment is made, we can construct a value function that treats the annuitization date T as given, and measures remaining utility at time t , from time t rather than from time zero, mapping into M-Y (2003, equation (8)),

$$V(W, t; T) \equiv \sup_{C_v, \Pi_v} E \left[\int_t^T e^{-r(v-t)} ({}_v p_{x+t}^s)^s \frac{(C_v - \hat{C})^{1-\gamma}}{1-\gamma} dv + e^{-r(T-t)} {}_{T-t} p_{x+t}^s \bar{a}_{x+T}^s \frac{\left(\frac{W_T}{\bar{a}_{x+T}^s} - \hat{C}\right)^{1-\gamma}}{1-\gamma} \mid W_t = W, \hat{C} \geq 0 \right] \tag{7}$$

where E denotes expectations. The difference between this problem and its counterpart M-Y (2003) is the presence here of a consumption floor. The two problems can be linked by a suitable transformation of the state variable. Specifically, define 'surplus' wealth \tilde{W} as the difference between actual wealth W and 'floor' or 'escrowed' wealth \hat{W} , defined by

$$\begin{aligned} \hat{W} &\equiv \frac{\hat{C}}{r} (1 - e^{r(t-T)}) + \hat{C} \bar{a}_{x+T}^s e^{r(t-T)}, \\ \tilde{W} &\equiv W - \hat{W}. \end{aligned} \tag{8}$$

The first term on the right-hand side of equation (8) can be interpreted as a fund that protects floor consumption prior to annuitization. Likewise, the second term can be interpreted as a fund that protects floor consumption after annuitization. Define

'surplus' consumption as $\tilde{C} \equiv C - \hat{C}$. We can now write the following HJB equation in W and other variables, corresponding to equation (9) of M-Y (2003)

$$(r + \lambda_{x+i}^s)V = V_t + \max_{\Pi} \left[\frac{1}{2} \sigma^2 \Pi^2 V_{WW} + (\alpha - r) \Pi V_W + r W V_W \right] + \max_{C \geq 0} \left[-C V_W + \frac{\tilde{C}^{1-\gamma}}{1-\gamma} \right], \quad (9)$$

with the boundary condition

$$V(W, T; T) = \frac{1}{1-\gamma} \left(\frac{\tilde{W}_T}{\bar{a}_{x+T}^s} \right)^{1-\gamma} \bar{a}_{x+T}^s. \quad (10)$$

Find expressions for the optimal controls

$$C^* = \hat{C} + (V_W)^{-\frac{1}{\gamma}} \quad (11)$$

and

$$\Pi^* = -\frac{(\alpha - r)}{\sigma^2} \frac{V_W}{V_{WW}}, \quad (12)$$

and substitute (11) and (12) into equation (9). The result is a PDE to be solved for an explicit expression for $V(W, t; T)$

$$(r + \lambda_{x+i}^s)V = V_t + \frac{\gamma}{1-\gamma} V_W^{-1/\gamma} + (rW - \hat{C})V_W - \frac{1}{2} \frac{(\alpha - r)^2}{\sigma^2} \frac{(V_W)^2}{V_{WW}}. \quad (13)$$

We propose a solution to the PDE in (13) in the form

$$V(W, t; T) = \mathcal{G}(t) \frac{1}{1-\gamma} (W - \tilde{W})^{1-\gamma}. \quad (14)$$

Recalling the definition of \tilde{W} from (8) and noting that $\tilde{W} = W - \hat{W}$, allows us to write the partial derivatives of $V(W, t; T)$ as

$$V_W = \mathcal{G}(t) \tilde{W}^{-\gamma} \quad (15)$$

$$V_{WW} = -\gamma \mathcal{G}(t) \tilde{W}^{-\gamma-1} \quad (16)$$

$$V_t = \mathcal{G}(t) \tilde{W}^{-\gamma} (\hat{C} e^{r(t-T)} - r \hat{C} \bar{a}_{x+T}^s e^{r(t-T)}) + \mathcal{G}'(t) \frac{\tilde{W}^{1-\gamma}}{1-\gamma}. \quad (17)$$

Substituting (14-17) into equation (13) gives

$$(r + \lambda_{x+i}^s) \mathcal{G}(t) \frac{\tilde{W}^{1-\gamma}}{1-\gamma} = \mathcal{G}(t) \tilde{W}^{-\gamma} (\hat{C} e^{r(t-T)} - r \hat{C} \bar{a}_{x+T}^s e^{r(t-T)}) + \mathcal{G}'(t) \frac{\tilde{W}^{1-\gamma}}{1-\gamma} + \frac{\gamma}{1-\gamma} \left[\mathcal{G}(t) \tilde{W}^{-\gamma} \right]^{\frac{\gamma-1}{\gamma}} + (rW - \hat{C}) \mathcal{G}(t) \tilde{W}^{-\gamma} + \frac{(\alpha - r)^2}{2\sigma^2 \gamma} \mathcal{G}(t) \tilde{W}^{1-\gamma}$$

$$= r \mathcal{G}(t) \tilde{W}^{-\gamma} \left\{ W - \left[\frac{\hat{C}}{r} (1 - e^{r(t-T)}) + \hat{C} \bar{a}_{x+T}^s e^{r(t-T)} \right] \right\} + \mathcal{G}'(t) \frac{\tilde{W}^{1-\gamma}}{1-\gamma} + \frac{\gamma}{1-\gamma} \left[\mathcal{G}(t) \right]^{\frac{\gamma-1}{\gamma}} \tilde{W}^{1-\gamma} + \frac{(\alpha - r)^2}{2\sigma^2 \gamma} \mathcal{G}(t) \tilde{W}^{1-\gamma} = r \mathcal{G}(t) \tilde{W}^{1-\gamma} + \mathcal{G}'(t) \frac{\tilde{W}^{1-\gamma}}{1-\gamma} + \frac{\gamma}{1-\gamma} \left[\mathcal{G}(t) \right]^{\frac{\gamma-1}{\gamma}} \tilde{W}^{1-\gamma} + \frac{(\alpha - r)^2}{2\sigma^2 \gamma} \mathcal{G}(t) \tilde{W}^{1-\gamma}$$

and

$$\frac{(r + \lambda_{x+i}^s) \mathcal{G}(t)}{1-\gamma} = r \mathcal{G}(t) + \frac{\mathcal{G}'(t)}{1-\gamma} + \frac{\gamma}{1-\gamma} \left[\mathcal{G}(t) \right]^{\frac{\gamma-1}{\gamma}} + \frac{(\alpha - r)^2}{2\sigma^2 \gamma} \mathcal{G}(t). \quad (19)$$

Rearranging equation (19) gives an ODE in $\mathcal{G}(t)$

$$0 = \mathcal{G}'(t) + \gamma \left[-g(t) \mathcal{G}(t) + \left[\mathcal{G}(t) \right]^{\frac{\gamma-1}{\gamma}} \right], \quad (20)$$

where $g(t) \equiv \frac{1-\gamma}{\gamma} \left[\frac{(r + \lambda_{x+i}^s)}{1-\gamma} - r - \frac{(\alpha - r)^2}{2\sigma^2 \gamma} \right]$. Appendix A presents a verification that the solution to (20) is consistent with M-Y (2003, equation (8)) such that

$$\mathcal{G}(t) = \left[\left\{ \frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^s)^{\frac{1}{1-\gamma}}} \right\}^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+i}^s)^{\frac{1}{\gamma}} + \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-t p_{x+i}^s)^{\frac{1}{\gamma}} dv \right]^{\gamma},$$

and the solution to the PDE (13) is

$$V(W, t; T) = \frac{1}{1-\gamma} \tilde{W}^{1-\gamma} \left[\left\{ \frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^s)^{\frac{1}{1-\gamma}}} \right\}^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+i}^s)^{\frac{1}{\gamma}} + \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-t p_{x+i}^s)^{\frac{1}{\gamma}} dv \right]^{\gamma} \quad (21)$$

where $\delta \equiv r + \frac{(\alpha - r)^2}{2\sigma^2 \gamma}$. The term in square brackets we define as $z(t, T) \equiv \mathcal{G}(t)^{\frac{1}{\gamma}}$.

For a pre-committed date of annuitization T the value function (21) coincides with the 'true' value function. Hence (21) can be differentiated to find the best approximate annuitization date. Appendix B shows that, if we calculate the derivative of

$V(W, t; T)$ in (21) with respect to T , we get:

$$\begin{aligned} \frac{\partial V}{\partial T} = & \frac{1}{1-\gamma} \tilde{W}^{1-\gamma} \gamma z(t, T)^{\gamma-1} \frac{1-\gamma}{\gamma} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} \left(T-tP_{x+t}^s \right)^{\frac{1}{\gamma}} \\ & \times \left[\frac{\gamma}{1-\gamma} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\left(\frac{1-\gamma}{\gamma} \right)} - \frac{1}{1-\gamma} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \bar{a}_{x+T}^s (\delta - (r + \lambda_{x+T}^o)) \right] \\ & - z(t, T)^\gamma \tilde{W}^{-\gamma} \hat{C} e^{-r(T-t)} \lambda_{x+T}^o \bar{a}_{x+T}^o. \end{aligned} \tag{22}$$

Equation (22) differs from its counterpart in M–Y (16) by the additional term

$$-z(t, T)^\gamma \tilde{W}^{-\gamma} \hat{C} e^{-r(T-t)} \lambda_{x+T}^o \bar{a}_{x+T}^o,$$

which is negative, and therefore a factor bringing forward annuitization.

The intuition behind (22) is straightforward. The HARA retiree protects all future subsistence consumption in escrow wealth, and consequently holds a smaller proportion of total wealth in the equity portfolio than the CRRA agent with $\hat{C}=0$. Lower exposure to the potentially high-yielding risky asset therefore reduces the option value of delaying annuitization. It follows that introducing a positive consumption floor has a similar effect to raising relative risk aversion. In addition, the agent recognizes that it is ‘cheaper’ to store escrow wealth in an annuity rather than a bond portfolio (at least where there are small enough loadings), creating another incentive to switch into complete annuitization at an earlier date.

Evaluating at $t=T$ and allowing subjective and objective survival probabilities to coincide simplifies (22) to

$$\left. \frac{\partial V}{\partial T} \right|_{t=T}^{s=0} = \tilde{W}_T^{1-\gamma} \bar{a}_{x+T}^\gamma [\delta - (r + \lambda_{x+T})] - \tilde{W}_T^{-\gamma} \bar{a}_{x+T}^\gamma \hat{C} \bar{a}_{x+T} \lambda_{x+T}. \tag{23}$$

Setting this expression equal to zero, and using the fact that $\hat{C} \bar{a}_{x+T} = \hat{W}_T$, shows that the optimal annuitization date will be decided by

$$\left. \frac{\partial V}{\partial T} \right|_{t=T}^{s=0} = 0 \implies \delta - r = \lambda_{x+T} \left(1 + \frac{\hat{W}_T}{\bar{W}_T} \right). \tag{24}$$

Notice firstly that under the M–Y assumption of no consumption floor ($\frac{\hat{W}_T}{\bar{W}_T} = 0$), the retiree’s optimal stopping problem in the $s=0$ case is a simple comparison between the risk-adjusted excess return to stocks and the return to annuities. Secondly, inclusion of a consumption floor brings the level of wealth into the solution to the retiree’s annuitization problem. This changes the nature of the solution, from being deterministic to approximate. In the absence of a consumption floor the retiree theoretically can know her date of annuitization decades in advance. In our case she must make an expectation of future wealth when evaluating alternative annuitization dates, and choose T according to (22) to maximize expected utility. Thirdly, a first-order condition characterizing T in terms of information available at any prior time t (including $t=0$) is derived in Appendix C.

Table 1. Approximate optimal age at annuitization

	Male (Female)		
	RRA	Sharpe ratio	
		0.18	0.30
Curvature parameter, $\gamma=0.5$			
Zero insured consumption	0.5	77.3 (81.6)	87.7 (90.7)
50 % insured consumption	1	70.6 (75.7)	80.9 (84.6)
Curvature parameter, $\gamma=1$			
Zero insured consumption	1	70.6 (75.7)	80.9 (84.6)
50 % insured consumption	2	63.8 (69.9)	74.2 (78.8)
Curvature parameter, $\gamma=2$			
Zero insured consumption	2	63.8 (69.9)	74.2 (78.8)
50 % insured consumption	4	57.0 (64.2)	67.4 (73.0)

One remarkable theoretical result of the Milevsky–Young model concerns the importance of divergent perceptions of the force of mortality to the optimal annuitization delay. In the HARA case, the additional weighting on the force of mortality evident in the right-hand term of (24) brings forward annuitization relative to the CRRA case, but the impact of divergent opinions on λ_{x+T} is similar. The proof outlined in Appendix D demonstrates that, whenever the individual thinks they are less likely to survive so that $\bar{a}_{x+T}^s < \bar{a}_{x+T}^o$, or when the individual thinks they are more healthy than average, in such a way that $\bar{a}_{x+T}^o < \bar{a}_{x+T}^s < 2 \bar{a}_{x+T}^o$, the optimal time to annuitize will be later than the T given by (22), even for a positive consumption floor. However the important caveat in the HARA case is that the presence of stochastic wealth \tilde{W} in the optimality condition makes all such comparisons approximate prior to the actual annuitization date.

To illustrate these ideas, the next section compares optimal annuitization timing for zero and insured consumption over a variety of risk tolerances and asset returns.

4 Numeric implications

One way to assess the impact of consumption insurance on annuitization is to apply observed mortality data and returns data to the model of Section 3. Once tastes for risk and subsistence are fixed, the optimal time to annuitize from (24) depends on comparison between the gains to risky asset exposure (here measured by δ and determined by the Sharpe ratio), and the value of the force of mortality scaled up by $\left(1 + \frac{\hat{W}_T}{\bar{W}_T} \right)$. Table 1 below shows the effects of increasing insured consumption from 0 to 50 per cent of total consumption. Fixing the proportion of subsistence consumption allows calculation of the relative risk aversion of the HARA agent, which, as noted earlier, is not constant, but falls as consumption increases above the floor. Recall also that the ratio $\frac{\hat{W}_T}{\bar{W}_T}$ appears in the optimality condition. This fraction approaches unity ($\frac{\hat{W}_T}{\bar{W}_T} = 1$) in the 50 per cent insured case at the point of annuitization.

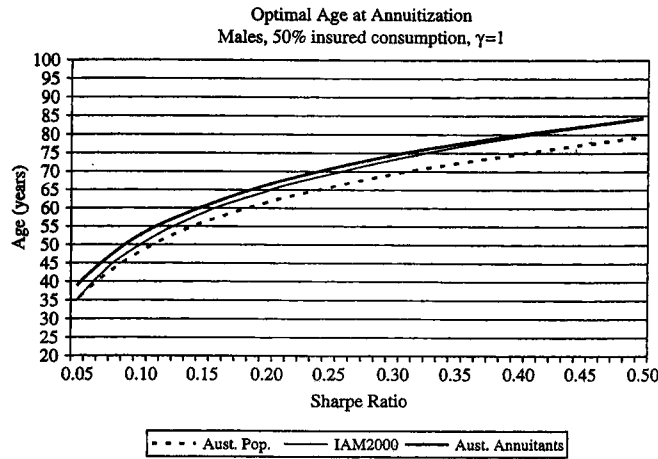


Figure 2. Optimal age at annuitization as Sharpe ratio varies

The Sharpe ratios of 0.18 and 0.30 underlying Table 1 obtain for two alternative forecasts of stock market returns, namely, $\alpha=0.06$, $r=0.03$, and $\sigma=0.17$, roughly in line with the views of Campbell (2002), or $\alpha=0.12$, $r=0.06$, and $\sigma=0.02$, consistent with M–Y and in line with the more optimistic views of Ibbotson (2002). Dispersion and modal parameters of the Gompertz distribution were estimated over Australian data⁸ for males (females) at $b=9.78(8.35)$ and $m=88.95(92.76)$.

There are two key points to make here. Firstly, the combination of a conservative forward-looking Sharpe ratio and a 50 per cent consumption floor causes any advantage in delayed annuitization to vanish for males and to shrink to about five years for females when $\gamma=1$. Delays are still preferred by the more risk tolerant. Not so, however, if choices are guided by an optimistic ‘historical’ assumption for the Sharpe ratio, linked to the high returns to equity that were recorded during the twentieth century in the United States, Australia, and a handful of other countries (Jorion and Goetzmann, 1999). Following Milevsky and Young, and using the higher Sharpe ratio raises the optimal delays to almost ten years for men and 14 years for women. Secondly, optimistic estimates of survival probability will also delay annuitization.

To gauge the importance of the Sharpe ratio and choice of mortality parameters to the delay, consider Figure 2. Figure 2 graphs the optimal annuitization date for a male agent with 50 per cent insured consumption across a range of Sharpe ratios and setting $\gamma=1$. Only as the Sharpe ratio rises above 0.20 does any advantage emerge in

⁸ The Gompertz function is used as a continuous approximation to discrete mortality tables. Parameters here were estimated as $\log(p_x) = \exp(\frac{ax}{b}) (1 - \exp(\frac{ax}{b}))$. Mortality data were from Australian Life Tables 95–97, using improved mortality discounted by 60%, to mimic the longevity of self-selecting annuitants. Improvements were calculated using the ABS method outlined in the Life Tables. For discussion of estimation methods generally see Valdez (2000) and Carriere (1992). For discussion of Australian practice in estimating annuitants’ mortality see Knox (2000) and Doyle, Mitchell and Piggott (2002).

Table 2. Age at annuitization, divergent survival rates, optimistic Sharpe ratio (Male)

Curvature parameter	Sharpe ratio = 0.30					
	$\gamma = 0.5$		$\gamma = 1$		$\gamma = 2$	
	50%	zero	50%	zero	50%	zero
Consumption floor						
factor f						
–0.8	83.71	95.18	75.54	84.36	68.10	75.64
–0.6	82.21	90.76	74.77	82.37	67.88	74.80
–0.4	81.43	88.85	74.39	81.47	67.48	74.39
–0.2	81.04	87.97	74.20	81.05	67.39	74.20
0	80.92	87.71	74.15	80.93	67.37	74.15
0.2	81.04	87.95	74.19	81.03	67.39	74.19
0.5	81.59	89.33	74.41	81.54	67.48	74.41
1.0	83.86	–	75.14	83.44	67.78	75.10
1.5	86.06	–	76.35	87.88	68.21	76.20
2.0	88.70	–	78.29	–	68.78	77.84

Note: A dash indicates that no stationary point of the value function existed in the range permitted by the parameterization.

delaying annuitization. Divergence in mortality estimates may account for around a 4–5 year variation in the optimal delay.

One key feature of the Milevsky and Young (2003) model is the potential for divergence between subjective and objective survival probabilities. Their extension of the standard single-distribution set-up is motivated by recent empirical evidence that subjective assessments of longevity have predictive power beyond the usual markers for length of life (such as smoking, disease etc.), and that individuals will revise subjective survival probabilities in response to new information (Hurd and McGarry, 2002). Including subjective views on mortality in models of annuitization can also help clarify the discussion of adverse selection, which has long been flagged as a reason for low rates of voluntary annuitization.

Results from the M–Y model show that when the decision to annuitize requires the complete and final transfer of all wealth, any disagreement about survival probabilities between the consumer and annuity provider motivates a delay: consumers in poor health find the terms of the annuity contract unfavorable; healthy consumers discount future consumption more slowly and hope for gains from longer exposure to the stock market. The numerical examples reported below (following M–Y) use the proportional hazards transformation to distinguish subjective and objective survival probabilities, where $\lambda_x^s = (1+f)\lambda_x^o$. A positive value of f means that the individual is less healthy than suggested by the objective standard, and negative values of f indicate that the individual is more healthy, with $-1 < f < \infty$.

Tables 2 and 3 present numerical estimates of the optimal time to annuitize for an individual who at $t=T$ insures 50 per cent of their consumption, compared with the

Table 3. Age at annuitization, divergent survival rates, conservative Sharpe ratio (Male)

Curvature parameter	Sharpe ratio = 0.18					
	$\gamma = 0.5$		$\gamma = 1$		$\gamma = 2$	
	50 %	zero	50 %	zero	50 %	zero
factor f						
-0.8	73.98	85.43	65.70	74.64	58.03	65.82
-0.6	72.05	80.55	64.57	72.20	57.42	64.61
-0.4	71.12	78.51	64.07	71.15	57.14	64.07
-0.2	70.67	77.59	63.82	70.68	57.02	63.83
0	70.55	77.33	63.77	70.55	56.99	63.77
0.2	70.66	77.57	63.82	70.66	57.01	63.82
0.5	71.22	78.85	64.70	71.19	57.14	64.07
1.0	73.35	86.71	64.88	73.03	57.51	64.83
1.5	78.96	-	66.15	76.81	58.03	66.01
2.0	80.90	-	68.02	-	58.70	67.66

case of zero consumption insurance. The subjective hazard rate is varied by the proportional transform f , from more to less healthy. The data in these tables show that asymmetries in mortality probabilities can still account for delays to annuitization. As well as bringing forward annuitization by at least 5–6 years, the consumption floor tends to compress the range of divergence. For the case where the Sharpe ratio is 0.3 and $\gamma = 2$, the range of optimal ages rises from 67.37 ($f = 0$) to only 68.78 ($f = 2$) for males with 50 per cent of consumption insured. The comparable range for zero insured consumption is much wider, rising from 74.15 to 77.85 years of age.

A less optimistic Sharpe ratio again reduces the advantages to delaying annuitization when subjective and objective survival estimates diverge. Results reported in Table 3 indicate earlier annuitization dates than Table 2, with moderately risk averse agents annuitizing before age 70, despite differing expectations of lifetime.

5 Conclusion

The analysis presented here extends a new explanation for the well-documented reluctance of retirees to purchase life annuities. As more of the developed world moves toward defined contribution retirement savings schemes, and more responsibility for the management of retirement incomes falls to the individual, the annuity puzzle becomes more pressing. Since the purchase of life annuities is irreversible, a real option to delay annuitization exists for any risk averse investors who enjoy the possibility of stochastic improvements to their budget constraint through ongoing investment in risky asset markets. Milevsky and Young (2002, 2003) describe and quantify this real option for investors with CRRA preferences. By extending the results to agents with HARA preferences we isolate the impact of fixed consumption

insurance, the simplest form of a habit persistence model, on the timing of optimal annuitization.

Two results are worth noting. Firstly, the desire to keep consumption above a specified floor creates an incentive to annuitize earlier than otherwise. HARA agents must maintain an escrow fund in the risk-free asset to cover future subsistence, effectively shrinking the potential for wealth creation through risky asset investment compared with CRRA agents, and making actuarially fair annuities more attractive. Secondly, divergence between a retiree's subjective assessment of their survival prospects and the annuity provider's objective assessment of their prospects will still add to any delay, as Milevsky and Young established.

Numerical estimates of the optimal annuitization date for a 65-year old male with a 50 per cent insured consumption floor depend on risk tolerance and forecasts of asset returns. For a plausible range of parameters, there is no advantage in delay. Putting off full annuitization will be better for females, for the more risk tolerant, for individuals who have optimistic expectation of investment returns, and for those who do not agree with market mortality estimates.

References

- Ameriks, J. and S. P. Zeldes (2001) How do portfolio shares vary with age? Unpublished manuscript, TIAA-CREF Institute, New York, NY.
- Australian Government Actuary (1999) Australian Life Tables 1995–97. Commonwealth of Australia, Parkes, ACT.
- Campbell, J. Y. (2002) Stock returns for a new century. TIAA-CREF Investment Forum (June): 11–12.
- Campbell, J. Y. and J. H. Cochrane (1999) By force of habit: a consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107 (April): 205–251.
- Carriere, J. F. (1992) Parametric models for life tables. Transactions, Society of Actuaries, Schaumburg IL., 44: 77–99.
- Constantinides, G. (1990) Habit formation: a resolution of the equity premium puzzle. *Journal of Political Economy*, 98 (June): 519–543.
- Davidoff, T., J. R. Brown and P. A. Diamond (2003) Annuities and individual welfare. NBER Working Paper no. w9714, National Bureau of Economic Research, Cambridge Massachusetts (May).
- Doyle, S., O. S. Mitchell, and J. Piggott (2002) Annuity values in defined contribution retirement systems: the case of Singapore and Australia. UNSW Centre for Pension and Superannuation Discussion Paper no. 02/02, University of New South Wales, Sydney.
- Hurd, M. D. and K. McGarry (2002) The predictive validity of subjective probabilities of survival. *Economic Journal*, 112 (October): 966–985.
- Ibbotson, R. G. (2002) Building the future from the past. TIAA-CREF Investment Forum (June): 10–12.
- Ingersoll, J. E. (1987) *Theory of Financial Decision Making*. Rowman & Littlefield, Savage.
- IFSA (2002) Retirement income streams report. Investment and Financial Services Association, Sydney, NSW (December).
- Jorion, P. and W. Goetzmann (1999) Global stock markets in the twentieth century. *Journal of Finance*, 54 (June): 953–980.
- Karatzas, I. and H. Wang (2000) Utility maximization with discretionary stopping. *SIAM Journal on Control and Optimization*, 39: 306–329.
- Kingston, G. (1989) Theoretical foundations of constant proportion portfolio insurance. *Economics Letters*, 29 (October): 345–347.

Knox, D. M. (2000) The Australian annuity market. World Bank Working Paper no. 2495. Development Research Group, World Bank, Washington, DC.

Leland, H. E. (1980) Who should buy portfolio insurance? *Journal of Finance*, **35** (May): 581–594.

Liu, J. and E. Neiss (2002) Endogenous retirement and portfolio choice. Working Paper, Anderson School of Business, UCLA, Los Angeles.

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

Merton, R. C. (1969) Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics*, **51** (August): 239–246.

Milevsky, M. A. and V. R. Young (2003) Annuitization and asset allocation. IFID Centre Working Paper, http://www.ifid.ca/pdf_workingpapers/WP2002B.pdf.

Milevsky, M. A. and V. R. Young (2002) Optimal asset allocation and the real option to delay annuitization: it's not now or never. IFID Centre Working Paper, http://www.ifid.ca/pdf_workingpapers/WP2003OCT15.pdf.

Mitchell, O. S. and D. McCarthy (2002) Annuities for an aging world. NBER Working Paper no. 9092, National Bureau of Economic Research, Cambridge Massachusetts (August).

Perold, A. F. and W. F. Sharpe (1988) Dynamic strategies for asset allocation. *Financial Analysts Journal* (January–February): 16–27.

Plan For Life (2003) The Pension and Annuity Market Research Report. Issue 39, May, Plan For Life Actuaries and Researchers, Mt. Waverly.

Rubinstein, M. (1976) The strong case for the generalized logarithmic utility model as the premier model of financial markets. *Journal of Finance*, **31** (May): 551–571.

Valdez, E. A. (2000) Developing a general law of mortality for Singapore. *Singapore International Insurance and Actuarial Journal* (June): 1–17.

Yaari, M. E. (1965) Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, **32** (April): 137–150.

Appendix A: Verification of ODE solution

Equation (20) is a Bernoulli ODE of the form

$$\mathcal{G}'(t) - \gamma g(t)\mathcal{G}(t) = -\gamma [\mathcal{G}(t)]^{\frac{\gamma-1}{\gamma}}, \tag{A.1}$$

with boundary condition

$$\mathcal{G}(T) = \frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}}. \tag{A.2}$$

Define $z(t, T) \equiv [\mathcal{G}(t)]^{\frac{1}{\gamma}}$ and rewrite (A.1) as:

$$dz + [-gz + 1]dt = 0$$

that is,

$$\frac{dz}{dt} = g(t)z(t, T) - 1, \tag{A.3}$$

with boundary condition

$$z(T) = \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{1/\gamma}. \tag{A.4}$$

M–Y propose a solution

$$z(t, T) = \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-tP_{x+t}^s)^{\frac{1}{\gamma}} + \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv. \tag{A.5}$$

By showing that the derivative of function (A.5) with respect to t satisfies equation (A.3), and showing that it meets the boundary condition (A.4), we verify that it is a solution to the ODE (A.1).

$$\begin{aligned} dz/dt &= \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} \frac{d}{dt} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-tP_{x+t}^s)^{\frac{1}{\gamma}} \\ &\quad + \frac{d}{dt} \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv. \end{aligned}$$

We can write, using Leibnitz's rule

$$\begin{aligned} \frac{d}{dt} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-tP_{x+t}^s)^{\frac{1}{\gamma}} &= \left(\frac{r-\delta(1-\gamma)}{\gamma} - \frac{d}{dt} \frac{1}{\gamma} \int_t^T \lambda_{x+v}^s dv \right) \\ &\quad \times e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-tP_{x+t}^s)^{\frac{1}{\gamma}} \\ &= \left(\frac{r-\delta(1-\gamma)}{\gamma} + \frac{1}{\gamma} \lambda_{x+t}^s \right) e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-tP_{x+t}^s)^{\frac{1}{\gamma}}. \end{aligned} \tag{A.6}$$

and

$$\begin{aligned} \frac{d}{dt} \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv &= -e^0 + \int_t^T \frac{\delta}{\delta t} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv \\ &= -1 + \int_t^T \left(\frac{r-\delta(1-\gamma)}{\gamma} - \frac{d}{dt} \frac{1}{\gamma} \int_t^v \lambda_{x+\eta}^s d\eta \right) \\ &\quad \times e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv \\ &= -1 + \int_t^T \left(\frac{r-\delta(1-\gamma)}{\gamma} + \frac{1}{\gamma} \lambda_{x+t}^s \right) \\ &\quad \times e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv \\ &= -1 + \left(\frac{r-\delta(1-\gamma)}{\gamma} + \frac{1}{\gamma} \lambda_{x+t}^s \right) \\ &\quad \times \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-tP_{x+t}^s)^{\frac{1}{\gamma}} dv. \end{aligned} \tag{A.7}$$

Combining terms gives

$$\begin{aligned} dz/dt &= \left(\frac{r-\delta(1-\gamma)}{\gamma} + \frac{1}{\gamma} \lambda_{x+t}^s \right) \left[\left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} \right. \\ &\quad \left. \times e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} + \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-t p_{x+t}^s)^{\frac{1}{\gamma}} dv \right] - 1 \\ &= g(t)z(t, T) - 1, \end{aligned} \quad (\text{A.8})$$

which verifies equation (A.3).

Finally we note that as $t \rightarrow T$

$$z(t, T) \rightarrow z(T) = \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}}, \quad (\text{A.9})$$

since $e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} \rightarrow 1$ and $\int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-t p_{x+t}^s)^{\frac{1}{\gamma}} dv \rightarrow 0$, and the boundary condition holds.

Appendix B: Derivation of equation (24)

Specify the value function

$$V(W, t; T) = \mathcal{G}(t) \frac{\bar{W}^{1-\gamma}}{1-\gamma} = z(t, T)^\gamma \frac{\bar{W}^{1-\gamma}}{1-\gamma} \quad (\text{B.1})$$

$$\begin{aligned} z(t, T)^\gamma &\equiv \left\{ \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} \right. \\ &\quad \left. + \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-t p_{x+t}^s)^{\frac{1}{\gamma}} dv \right\}^\gamma \end{aligned} \quad (\text{B.2})$$

$$\hat{W}_t \equiv \frac{\hat{C}}{r} (1 - e^{r(t-T)}) + \hat{C} \bar{a}_{x+T}^o e^{r(t-T)}. \quad (\text{B.3})$$

Differentiating this function with respect to T gives the optimal time to annuitize. Using the product and chain rules,

$$\frac{\partial V}{\partial T} = \frac{\bar{W}^{1-\gamma}}{1-\gamma} \gamma z(t, T)^{\gamma-1} \frac{\partial z(t, T)}{\partial T} + z(t, T)^\gamma \bar{W}^{-\gamma} \frac{\partial \bar{W}}{\partial T}. \quad (\text{B.4})$$

The first term in (B.4) is given by $\frac{\bar{W}^{1-\gamma}}{1-\gamma} \gamma z(t, T)^{\gamma-1} \frac{\partial z(t, T)}{\partial T}$, where:

$$\frac{\partial z(t, T)}{\partial T} = \frac{\partial}{\partial T} \left\{ \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} \right. \\ \left. + \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(v-t)} (v-t p_{x+t}^s)^{\frac{1}{\gamma}} dv \right\}.$$

We note that $z(t, T)$ has two additive terms. Consider the first term:

$$\left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}},$$

where

$$\begin{aligned} &\frac{\partial}{\partial T} \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} \\ &= \frac{1}{\gamma} \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} \left\{ (\lambda_{x+T}^s + r) - \frac{1}{\bar{a}_{x+T}^s} - (1-\gamma) \left[(\lambda_{x+T}^o + r) - \frac{1}{\bar{a}_{x+T}^o} \right] \right\}. \end{aligned}$$

In addition

$$\frac{\partial}{\partial T} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} = -\frac{r-\delta(1-\gamma)}{\gamma} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)},$$

and

$$\frac{\partial}{\partial T} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} = \frac{-\lambda_{x+T}^s}{\gamma} (T-t p_{x+t}^s)^{\frac{1}{\gamma}}$$

by Leibnitz's rule.

Together

$$\begin{aligned} \frac{\partial}{\partial T} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} &= \frac{1}{\gamma} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} \\ &\quad \times [(-\lambda_{x+T}^s) - (r-\delta(1-\gamma))]. \end{aligned}$$

Then using the product rule again

$$\begin{aligned} &\frac{\partial}{\partial T} \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} \\ &= \frac{1}{\gamma} \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}} \\ &\quad \times \left[(1-\gamma)(\delta - (\lambda_{x+T}^o + r)) - \frac{1}{\bar{a}_{x+T}^s} + \frac{(1-\gamma)}{\bar{a}_{x+T}^o} \right]. \end{aligned} \quad (\text{B.5})$$

Now differentiate the second additive term in $z(t, T)$ by Leibnitz's rule to get

$$\frac{\partial}{\partial T} \int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(s-t)} (s-t p_{x+t}^s)^{\frac{1}{\gamma}} ds = e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t p_{x+t}^s)^{\frac{1}{\gamma}}. \quad (\text{B.6})$$

The sum of (B.5) and (B.6) shows that

$$\frac{\partial z(t, T)}{\partial T} = \frac{(1-\gamma)}{\gamma} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t)^{\frac{1}{\gamma}} \times \left[\bar{a}_{x+T}^s (\delta - (\lambda_{x+T}^o + r)) - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\frac{1-\gamma}{\gamma}} \right].$$

Thus the first term in (B.4) is given by

$$\frac{\bar{W}^{1-\gamma}}{1-\gamma} \gamma z(t, T)^{\gamma-1} \frac{\partial z(t, T)}{\partial T} = \frac{\bar{W}^{1-\gamma}}{1-\gamma} \gamma z(t, T)^{\gamma-1} \times e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t)^{\frac{1}{\gamma}} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} \times \left\{ \bar{a}_{x+T}^s (\delta - \lambda_{x+T}^o - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\frac{1-\gamma}{\gamma}} \right\}. \tag{B.7}$$

This expression is consistent with M-Y 2003 equation (16), where the authors note that without a consumption floor,

$$\frac{\partial V}{\partial T} \propto \bar{a}_{x+T}^s (\delta - \lambda_{x+T}^o - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\frac{1-\gamma}{\gamma}}. \tag{B.8}$$

However, with a consumption floor the derivative has another additive term,

$$z(t, T)^\gamma \bar{W}^{-\gamma} \frac{\partial \bar{W}}{\partial T}. \tag{B.9}$$

Now

$$\begin{aligned} \frac{\partial \bar{W}}{\partial T} &= \frac{\partial(-\hat{W})}{\partial T} \\ &= - \left[\frac{\partial \hat{C}(1 - e^{r(t-T)})}{\partial T} + \frac{\partial \hat{C} \bar{a}_{x+T}^o e^{r(t-T)}}{\partial T} \right] \\ &= - \hat{C} e^{r(t-T)} \lambda_{x+T}^o \bar{a}_{x+T}^o. \end{aligned} \tag{B.10}$$

Combining terms from (B.7), B.9, and B.10 produces a complete expression for the derivative:

$$\frac{\partial V}{\partial T} = \frac{\bar{W}^{1-\gamma}}{1-\gamma} \gamma z(t, T)^{\gamma-1}$$

$$\begin{aligned} &\times e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} (T-t)^{\frac{1}{\gamma}} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} \\ &\times \left\{ \bar{a}_{x+T}^s (\delta - \lambda_{x+T}^o - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\frac{1-\gamma}{\gamma}} \right\} \\ &- z(t, T)^\gamma \bar{W}^{-\gamma} \hat{C} e^{r(t-T)} \lambda_{x+T}^o \bar{a}_{x+T}^o. \end{aligned} \tag{B.11}$$

By setting $t = T$, note that

$$z(T) = \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{1}{\gamma}}, \tag{B.12}$$

hence,

$$\begin{aligned} \frac{\partial V}{\partial T} \Big|_{t=T} &= \frac{\bar{W}_T^{1-\gamma}}{1-\gamma} \gamma \left(\frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \right)^{\frac{\gamma-1}{\gamma}} \times \frac{(1-\gamma)}{\gamma} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} \\ &\times \left\{ \bar{a}_{x+T}^s (\delta - \lambda_{x+T}^o - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\frac{1-\gamma}{\gamma}} \right\} \\ &- \frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \bar{W}_T^{-\gamma} [\hat{C} \lambda_{x+T}^o \bar{a}_{x+T}^o] \\ &= \bar{W}_T^{-\gamma} \left[\frac{\bar{W}_T}{(\bar{a}_{x+T}^o)^{1-\gamma}} \left\{ \bar{a}_{x+T}^s (\delta - \lambda_{x+T}^o - r) - \frac{1}{(1-\gamma)} \right. \right. \\ &\left. \left. + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{-\frac{1-\gamma}{\gamma}} \right\} - \frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \hat{C} \lambda_{x+T}^o \bar{a}_{x+T}^o \right], \end{aligned} \tag{B.13}$$

and if $s = o$

$$\begin{aligned} \frac{\partial V}{\partial T} \Big|_{t=T}^{s=o} &= \bar{W}_T^{-\gamma} (\bar{a}_{x+T})^\gamma [\bar{W}_T (\delta - \lambda_{x+T} - r) - \hat{C} \lambda_{x+T} \bar{a}_{x+T}] \\ \frac{\partial V}{\partial T} \Big|_{t=T}^{s=o} &\propto \bar{W}_T (\delta - \lambda_{x+T} - r) - \hat{C} \lambda_{x+T} \bar{a}_{x+T} \\ \frac{\partial V}{\partial T} \Big|_{t=T}^{s=o} &\propto (\delta - r) - \lambda_{x+T} \left(\frac{\hat{W}_T}{\bar{W}_T} + 1 \right). \end{aligned} \tag{B.14}$$

Appendix C: An approximate solution to the pre-annuitization problem

First-order condition

Assume for simplicity that $o=s$. Consider the value function

$$\bar{V}(\bar{W}, t) \equiv \sup_{C_t, \Pi_t, T} E_t \left[\int_t^T e^{-r(s-t)} p_{x+t} \frac{(C_s - \bar{C})^{1-\gamma}}{1-\gamma} ds + k(t, T) \frac{\bar{W}^{1-\gamma}}{1-\gamma} \mid \bar{W}_t = \bar{W} \right] \quad (C.1)$$

$$\text{where } k(t, T) \equiv e^{-r(T-t)} p_{x+t} \bar{a}_{x+T}^\gamma.$$

Expected net gain from committing to annuitize at date T , with the decision being taken at date t is zero if and only if

$$\left. \frac{\partial E_t[\dots]}{\partial T} \right|_{C_t, \Pi_t, \text{conditionally optimized}} = 0. \quad (C.2)$$

First order condition with respect to T is proportional to:

$$\begin{aligned} & E_t d\bar{W}_T^{1-\gamma} \Big|_{\text{commit at } t \text{ to annuitize at } T} \\ & \propto E_t \left[\bar{W}_T^{-\gamma} d\bar{W}_T - \frac{\gamma}{2} \bar{W}_T^{-(1+\gamma)} (d\bar{W}_T)^2 \right] \\ & = E_t \left\{ \bar{W}_T^{-\gamma} [\Pi_T(\alpha-r)dt + \sigma\Pi_T dz - W_T \lambda_{x+T} dt] - \frac{\gamma}{2} \bar{W}_T^{-(1+\gamma)} \sigma^2 \Pi_T^2 dt \right\}. \end{aligned} \quad (C.3)$$

Substitute for optimized Π_T to get

$$= E_t \left\{ \bar{W}_T^{-\gamma} \left[\frac{(\alpha-r)^2}{\gamma\sigma^2} \bar{W}_T dt + \frac{(\alpha-r)}{\gamma\sigma} \bar{W}_T dz - W_T \lambda_{x+T} dt \right] - \frac{1}{2} \bar{W}_T^{-(1+\gamma)} \frac{(\alpha-r)^2}{\gamma\sigma^2} \bar{W}_T^2 dt \right\}. \quad (C.4)$$

Collect terms

$$= E_t \left\{ \bar{W}_T^{1-\gamma} \left[\frac{(\alpha-r)^2}{\gamma\sigma^2} - \frac{1}{2} \frac{(\alpha-r)^2}{\gamma\sigma^2} \right] dt - \bar{W}_T^{-\gamma} W_T \lambda_{x+T} dt + \bar{W}_T^{1-\gamma} \frac{(\alpha-r)}{\gamma\sigma} dz \right\}, \quad (C.5)$$

run the expectations operator through

$$= \frac{1}{2} \frac{(\alpha-r)^2}{\gamma\sigma^2} E_t(\bar{W}_T^{1-\gamma}) dt - E_t[\bar{W}_T^{-\gamma} (\bar{W}_T + \hat{W}_T)] \lambda_{x+T} dt + 0 \quad (C.6)$$

$$= \frac{1}{2} \frac{(\alpha-r)^2}{\gamma\sigma^2} E_t(\bar{W}_T^{1-\gamma}) dt - \lambda_{x+T} [E_t(\bar{W}_T^{1-\gamma}) dt + \hat{W}_T E_t(\bar{W}_T^{-\gamma}) dt], \quad (C.7)$$

which will be zero when

$$\frac{1}{2} \frac{(\alpha-r)^2}{\gamma\sigma^2} = \lambda_{x+T} \left[1 + \frac{\hat{W}_T E_t(\bar{W}_T^{-\gamma})}{E_t(\bar{W}_T^{1-\gamma})} \right]. \quad (C.8)$$

In the limiting case $t=T$, this reduces to equation (24) of the main text.

Appendix D: Optimal T when subjective and objective hazard rates diverge

Milevsky and Young (2003) Appendix B presents a proof of the proposition that annuitization is delayed when subjective and objective assessments of the force of mortality are different, but obey $\bar{a}_{x+T}^s < 2\bar{a}_{x+T}^o$. In the case where an individual views themselves as less likely to survive, this condition is always met because $\bar{a}_{x+T}^s < \bar{a}_{x+T}^o$, but the proof also holds for individuals who regard themselves as more likely to survive, as long as $\bar{a}_{x+T}^s < 2\bar{a}_{x+T}^o$. (Numerical examples give more general support for the result.) In the case of HARA utility the divergence result still holds. The following adapts the M-Y proof to the HARA case.

The optimal moment of annuitization occurs when

$$\begin{aligned} \frac{\partial V}{\partial T} \Big|_{t=T} &= \bar{W}_T^{-\gamma} \left[\frac{\bar{W}_T}{(\bar{a}_{x+T}^o)^{1-\gamma}} \left\{ \bar{a}_{x+T}^s (\delta - \lambda_{x+T}^o - r) - \frac{1}{(1-\gamma)} \right. \right. \\ &\quad \left. \left. + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} \right\} - \frac{\bar{a}_{x+T}^s}{(\bar{a}_{x+T}^o)^{1-\gamma}} \hat{C} \lambda_{x+T} \bar{a}_{x+T}^o \right]. \end{aligned}$$

Setting this expression equal to zero gives

$$0 = \bar{a}_{x+T}^s \left(\delta - \lambda_{x+T}^o \left(1 + \frac{\bar{W}_T}{\bar{W}_T} \right) - r \right) + \left[\frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^s}{\bar{a}_{x+T}^o} \right]. \quad (D.1)$$

Define the subjective annuity factor in terms of the objective annuity factor as

$$\bar{a}_{x+T}^s = \bar{a}_{x+T}^o + \varepsilon \quad (D.2)$$

for small ε of either sign, and rewrite (D.1)

$$\begin{aligned} 0 &= (\bar{a}_{x+T}^o + \varepsilon) \left[\delta - \lambda_{x+T}^o \left(1 + \frac{\bar{W}_T}{\bar{W}_T} \right) - r \right] \\ &\quad + \left[\frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}_{x+T}^o + \varepsilon}{\bar{a}_{x+T}^o} \right)^{\frac{1-\gamma}{\gamma}} - \frac{1}{(1-\gamma)} + \frac{\bar{a}_{x+T}^o + \varepsilon}{\bar{a}_{x+T}^o} \right]. \end{aligned} \quad (D.3)$$

By expanding the second term around zero, this expression reduces to

$$\begin{aligned} 0 &= (\bar{a}_{x+T}^o + \varepsilon) \left[\delta - \lambda_{x+T}^o \left(1 + \frac{\bar{W}_T}{\bar{W}_T} \right) - r \right] - \frac{1}{2} \left(-\frac{1-\gamma}{\gamma} - 1 \right) \left(\frac{\varepsilon}{\bar{a}_{x+T}^o} \right)^2 \\ &\quad - \frac{1}{6} \left(-\frac{1-\gamma}{\gamma} - 1 \right) \left(-\frac{1-\gamma}{\gamma} - 2 \right) \left(\frac{\varepsilon}{\bar{a}_{x+T}^o} \right)^3 + \dots, \end{aligned} \quad (D.4)$$

for $-1 < \frac{\varepsilon}{\bar{a}_{x+T}^o} < 1$.

Milevsky and Young state that by choosing a value $0 < \varepsilon^* < \frac{\varepsilon}{\bar{a}_{x+T}^o}$, the mean value theorem gives

$$\begin{aligned} 0 &= (\bar{a}_{x+T}^o + \varepsilon) \left[\delta - \lambda_{x+T}^o \left(1 + \frac{\hat{W}_T}{\bar{W}_T} \right) - r \right] + \frac{(\varepsilon^*)^2}{2\gamma} \\ &= \left[\delta - \lambda_{x+T}^o \left(1 + \frac{\hat{W}_T}{\bar{W}_T} \right) - r \right] + \frac{(\varepsilon^*)^2}{2\gamma} \frac{1}{\bar{a}_{x+T}^o \left(1 + \frac{\varepsilon}{\bar{a}_{x+T}^o} \right)}. \end{aligned} \quad (\text{D.5})$$

Since $-1 < \frac{\varepsilon}{\bar{a}_{x+T}^o} < 1$, the second term is always positive. The condition for optimal annuitization is

$$\delta - r + \frac{(\varepsilon^*)^2}{2\gamma} \frac{1}{\bar{a}_{x+T}^o \left(1 + \frac{\varepsilon}{\bar{a}_{x+T}^o} \right)} = \lambda_{x+T}^o \left(1 + \frac{\hat{W}_T}{\bar{W}_T} \right).$$

So we can infer that differences between hazard rates still delay annuitization in the region $-1 < \frac{\varepsilon}{\bar{a}_{x+T}^o} < 1$, or equivalently, $\bar{a}_{x+T}^s < 2\bar{a}_{x+T}^o$.