Annuitization and Asset Allocation with HARA Utility

Geoffrey Kingston and Susan Thorp

School of Economics

University of New South Wales

Sydney

March 2004

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 annutization (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 of the RODA analysis to the case of HARA (or GLUM) preferences, the simplest representation of a consumption habit. The formula for the optimal timing of annuitization is surprisingly simple, but yields only a myopic solution, that is, the precise date of annuitization cannot be ascertained in advance. 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 forward-looking estimates of the Sharpe ratio.

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, and Davidoff, Brown and Diamond 2003). But despite the ready availability of longevity insurance retirees across the world seldom voluntarily annuitize.

Milevsky and Young (2002, 2003) have come up with a new explanation of the reluctance of retirees to buy life annuities. They begin with the observation that re-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 microeconomists 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.

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 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

[1]For evidence that the investement 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.

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. The difference is that the overall delay is reduced 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.

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 expen-

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

sive 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 mandated percentage of their earnings to retirement savings, no-one is compelled to purchase longevity insurance at retirement. A means-tested government pension acts as a safety-net to people over 65 years of age, and an array of tax-preferred income stream choices are on offer. These include immediate annuities (life and term) and phased withdrawal products termed '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 2003) and 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

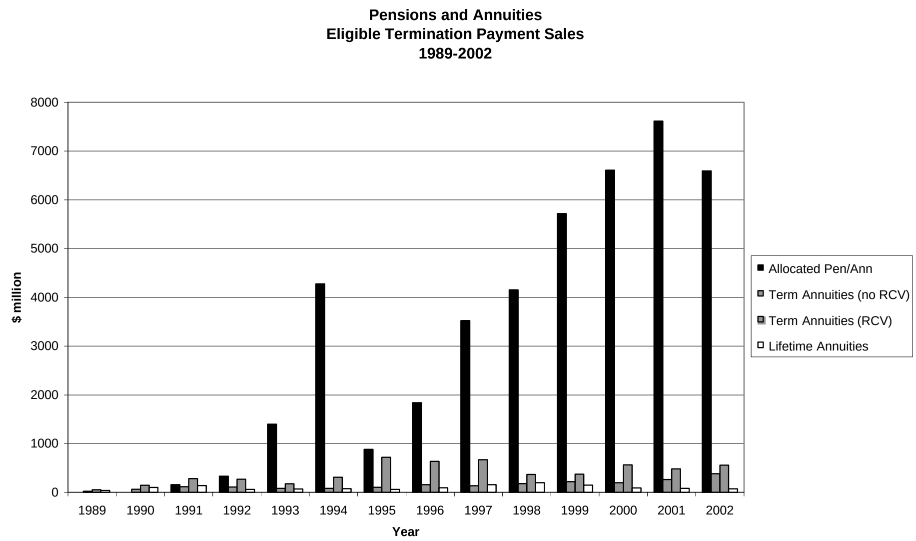
[4] Mitchell and McCarthy (2002) provide a detailed account of these issues.

[5] Few of these would have involved purchased products. (See Figure 1.)

[6] Australian regulators have recently flagged the introduction of market-linked income stream products termed 'growth accounts'. 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 to the proposed growth accounts.

annuity with no residual capital value, they are allowed a higher tax-concessional income, and the value of the annuity is not counted in social security means tests. Despite this, **Figure 1** below shows that people do not wish to purchase longevity insurance at current terms. 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, 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 is evidently appealing.

Figure 1: Australian Sales of Retirement Income Streams



Source: Plan For Life (2003)

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, has continued to decline in favour of term annuities. And furthermore the most popular term annuities are those which return residual capital. As noted above, 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) 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 which 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 an instantaneous 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 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^b_{x+T+t} ($t \geq 0$). In the case of time-separable HARA preferences, discounted direct utility at time t after annuitization is

$$U(C,t;T) = \frac{(C_t - \hat{C})^{1-\gamma}}{1-\gamma} e^{-(rt + \int_{x+T}^{x+T+t} \lambda^b_s ds)} \tag{1}$$

so that the corresponding equation for indirect utility is

$$\frac{(\frac{W_T}{\bar{a}^o_{x+T}} - \hat{C})^{1-\gamma}}{1-\gamma} e^{-(rt + \int_{x+T}^{x+T+t} \lambda^b_s ds)}, \tag{2}$$

[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).

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

$$\begin{aligned} \bar{a}_x &= \int_0^\infty e^{-rt}{}_t p_x dt \\ {}_t p_x &= e^{-\int_0^t \lambda_{x+s} ds}, \end{aligned} \tag{3}$$

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

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

$$V(W,0,T) \equiv e^{-rT}{}_T p^b_x \bar{a}^b_{x+T} \frac{(\frac{W}{\bar{a}^o_{x+T}} - \hat{C})^{1-\gamma}}{1-\gamma} \tag{4}$$

where $\bar{a}^b_{x+T}$ is the subjective annuity factor for an agent aged $x+T$, and $\bar{a}^o_{x+T}$ is the analogous factor between the objective probability of survival, ${}_t p^o_x$ and the objective hazard function λ^o_{x+t}. Equation (4) implies a boundary condition $V(W,T,T)$ [$= \bar{a}^b_{x+T} \frac{\left(\frac{W}{\bar{a}^O_{x+T}} - \hat{C}\right)^{1-\gamma}}{1-\gamma}$] 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 μ 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,

$$dS(t) = \mu S(t)dt + \sigma S(t)dz$$

where dz is a standard Wiener process. In the numerical analysis below values for μ 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_t invested in stocks, along with consumption C_t, and a date T on which the retiree's wealth is annuitized:

$$\max_{C_t,\ \Pi_t,\ T} E[\int_0^T e^{-rt}({}_tp_x^b)\frac{(C_t - \hat{C})}{1-\gamma}^{1-\gamma} dt \ + \ V(W,0,T)] \ , \tag{5}$$

subject to

$$\begin{aligned} dW_s &= [rW_s + (\alpha - r)\Pi_s - C_s]ds + \sigma\Pi_s dz_s \ , \qquad (6) \\ W_t &= w > 0. \end{aligned}$$

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

3.3. Solution

Following Milevsky and Young a step towards solving our problem is to 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 exactly into M-Y (2003) equation (8):

$$V(w,t;T) \equiv \sup_{C_s,\ \Pi_s} E[\int_t^T e^{-r(s-t)}({}_{s-t}p^b_{x+t})\frac{(C_s-\hat{C})^{1-\gamma}}{1-\gamma}ds \tag{7}$$

$$+ e^{-r(T-t)}{}_{T-t}p^b_{x+t}\bar{a}^b_{x+T}\frac{(\frac{W_T}{\bar{a}^o_{x+T}}-\hat{C})^{1-\gamma}}{1-\gamma} \quad |W_t=w, \quad \hat{C}\geqslant 0].$$

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 matched up by a suitable transformation of the state variable. Specifically, define 'surplus' wealth $\tilde{W}_t$ as the difference between actual wealth W_t and 'floor' or 'escrowed' wealth $\hat{W}_t$, defined as

$$\begin{aligned}\hat{W}_t &\equiv \frac{\hat{C}}{r}(1 - e^{r(t-T)}) + \hat{C}\bar{a}^o_{x+T}\ e^{r(t-T)}. \qquad (8)\\ \tilde{W}_t &\equiv W_t - \hat{W}_t\end{aligned}$$

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 $\tilde{C}_t \equiv C_t - \hat{C}_t$. Then the evolution of surplus wealth is related to surplus consumption via

$$\begin{aligned}d\tilde{w}_s &= [r\tilde{w}_s + (\alpha - r)\Pi_s - \tilde{C}_s]ds + \sigma\Pi_s dz_s \qquad (9)\\ \tilde{W}_t &= \tilde{w} > 0.\end{aligned}$$

This shows that our particular specification of 'surplus' wealth does the job of matching up the two problems under consideration. Consider the following HJB equation in $\tilde{w}$ and other variables, corresponding to equation (9) of M-Y 2003:

$$(r + \lambda^b_{x+t})V = V_t + \max_{\Pi}[\frac{1}{2}\sigma^2\Pi^2 V_{\tilde{w}\tilde{w}} + (\alpha - r)\Pi V_{\tilde{w}}] + r\tilde{w}V_{\tilde{w}} + \max_{\tilde{C}\geq 0}[-\tilde{C}V_{\tilde{w}} + \frac{\tilde{C}^{1-\gamma}}{1-\gamma}], \quad (10)$$

with the boundary condition

$$V(\tilde{w},T;T)=\frac{1}{1-\gamma}(\frac{\tilde{w}}{\bar{a}^o_{x+T}})^{1-\gamma}\bar{a}^b_{x+T}. \tag{11}$$

M-Y's equation (12) shows that the solution to this PDE is given by:

$$\begin{aligned} V(\tilde{w},t;T) \quad &= \quad \frac{1}{1-\gamma}\tilde{w}^{1-\gamma}\left[\left\{\frac{\bar{a}^b_{x+T}}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}}\right\}^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)}\left({}_{T-t}p^b_{x+t}\right)^{\frac{1}{\gamma}}\right. \\ &\left.+\int_t^T e^{-\frac{r-\delta(1-\gamma)}{\gamma}(s-t)}\left({}_{s-t}p^b_{x+t}\right)^{\frac{1}{\gamma}} ds.\right]^{\gamma} \end{aligned} \tag{12}$$

where $\delta \equiv r+\frac{(\mu-r)^2}{2\sigma^2\gamma}$.

At the date of annuitization T, the value function (12) coincides with the 'true' value function. Hence (12) can be differentiated to find the true annuitization date. Defining the term in square brackets in (12) above as $A(t)$,and calculating the derivative of V with respect to T, we get:

$$\begin{aligned} \frac{\partial V}{\partial T} \quad &= \quad \frac{1}{1-\gamma}\tilde{w}^{1-\gamma}\gamma A(t)^{\gamma-1}\frac{1-\gamma}{\gamma}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{\frac{1-\gamma}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)}\left({}_{T-t}p^b_{x+t}\right)^{\frac{1}{\gamma}} \text{ x} \\ &\left[\frac{\gamma}{1-\gamma}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{-\left(\frac{1-\gamma}{\gamma}\right)}-\frac{1}{1-\gamma}+\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}+\bar{a}^b_{x+T}(\delta-\left(r+\lambda^o_{x+T}\right))\right] \\ &-A(t)^{\gamma}\tilde{w}^{-\gamma}\left[\hat{C}e^{-r(T-t)}\lambda^o_{x+T}\bar{a}^o_{x+T}\right] \end{aligned} \tag{13}$$

Equation (13) differs from its counterpart in M-Y (16) by the additional term

$$-A(t)^{\gamma}\tilde{w}^{-\gamma}\left[\hat{C}e^{-r(T-t)}\lambda^{o}_{x+T}\bar{a}^{o}_{x+T}\right]$$

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

The intuition behind (13) is straightforward. The HARA retiree is already protecting 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 over an infinite horizon (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 (13) to:

$$\left.\frac{\partial V}{\partial T}\right|_{t=T}^{b=o} = \tilde{W}_T^{1-\gamma}\bar{a}^{\gamma}_{x+T}[\delta - (r + \lambda_{x+T})] - \tilde{W}_T^{-\gamma}\bar{a}^{\gamma}_{x+T}[\hat{C}\bar{a}_{x+T}\lambda_{x+T}] \tag{14}$$

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:

$$\left.\frac{\partial V}{\partial T}\right|_{t=T}^{b=o} = 0 \implies \delta - r = \lambda_{x+T}(1 + \frac{\hat{W}_T}{\tilde{W}_T}) \tag{15}$$

Notice firstly that under the M-Y assumption of no consumption floor ($\frac{\hat{W}_T}{\tilde{W}_T} = 0$), the retiree's optimal stopping problem in the $b = o$ 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 optimal stopping problem. This changes the nature of the solution, from being deterministic to myopic. In other words, you only know precisely when you are going to annuitize at the instant of doing so. In the absence of a consumption floor, by contrast, the retiree theoretically knows her date of annuitization decades in advance.[8] Finally, since the time of annuitization here is stochastic during almost all the pre-annuitisation phase, our value function (12) is exact (rather than approximate) only at time T. Hence, exact values of RODAs cannot be calculated in our setting.[9]

One remarkable theoretical result of the Milevsky-Young model was 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

[8]We hasten to add that incorporation of the real-world feature of a random element in the time variation of mortality hazard rates would constitute an equally reasonable way of changing the Milevsky-Young solution into a myopic one.

[9]Except in the unlikely event that an exact solution to the problem (7) can be found. Recall, however, that our solution to the optimal stopping problem is not subject to this limitation.

right hand term of (15) 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 B demonstrates that whenever the individual thinks they are less likely to survive so that $\bar{a}^b_{x+T} < \bar{a}^o_{x+T}$, or when the individual thinks they are more healthy than average, in such a way that $\bar{a}^b_{x+T} < 2\bar{a}^o_{x+T}$, the optimal time to annuitize will be later than the T given by (15) 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 (15) depends crucially 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 $(1 + \frac{\hat{W}_T}{\tilde{W}_T})$. **Table 1** below shows the effects of increasing insured consumption from zero to fifty 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 $\frac{\hat{W}_T}{\tilde{W}_T}$ appears in the optimality condition. This fraction approaches unity at the point of annuitization. The table below sets $\frac{\hat{W}_T}{\tilde{W}_T} = 1$ in the 50% insured case.

Table 1

Approximate Optimal Age at Annuitization

Male (Female)

		Sharpe Ratio	
Curvature parameter, $\gamma = 0.5$	**RRA**	**0.18**	**0.30**
Zero insured consumption	0.5	77.7 (81.9)	87.7 (90.4)
50% insured Consumption	1	71.6 (76.1)	81.6 (84.6)
Curvature parameter, $\gamma = 1$			
Zero insured consumption	1	70.9 (76.1)	80.9 (84.6)
50% insured Consumption	2	64.1 (70.3)	74.1 (78.8)
Curvature parameter, $\gamma = 2$			
Zero insured consumption	2	64.1 (70.3)	74.1 (78.8)
50% insured Consumption	4	57.4 (64.5)	67.4 (73.0)

The Sharpe ratios of .18 and .30 underlying Table 1 obtain for two alternative forecasts of stock market returns, namely, $\mu = .06$, $r = .03$, and $\sigma = .17$, roughly in line with the views of Campbell (2002) or $\mu = .12$, $r = .06$, and $\sigma = .20$, 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 data[10] for males (females) at $b = 9.78(8.35)$ and $m = 88.95(92.76)$.

There a 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 5 years for females when risk aversion is one.[11] 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 20th 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 10 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 **Figures 2-4**. These figures graph the optimal annuitization date for a male agent with 50 per cent insured consumption across a range of Sharpe ratios,

[10] The Gompertz function is used as a continuous approximation to discrete mortality tables. Parameters here were estimated as $\log(p_x) = \exp\left(\frac{x-m}{b}\right)\left(1 - \exp\frac{1}{b}\right)$. 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 annuitiants' mortality see Knox (2000) and Doyle, Mitchell and Piggott (2002).

[11] Quotes for a \$100,000, CPI-indexed immediate life annuity for a 65 year old female (with a 10 year guarantee) in Australia offer initial income of \$4892. Assuming a risk-free real interest rate of 3 per cent, this quote implies (from the purchaser's perspective) an average life expectancy for female annuitants of around 97 years, representing a discounting of population mortality estimates in the order of 55-60 per cent. This is consistent with the parameters underlying Table 1.

using three alternative mortality scenarios and the three chosen values for γ. For the case of $\gamma = 1$, for example, only as the Sharpe ratio rises above 0.20 does any advantage emerge in delaying annuitization. Divergence in mortality estimates may account for around a 4-5 year variation in the optimal delay.

Figure 2

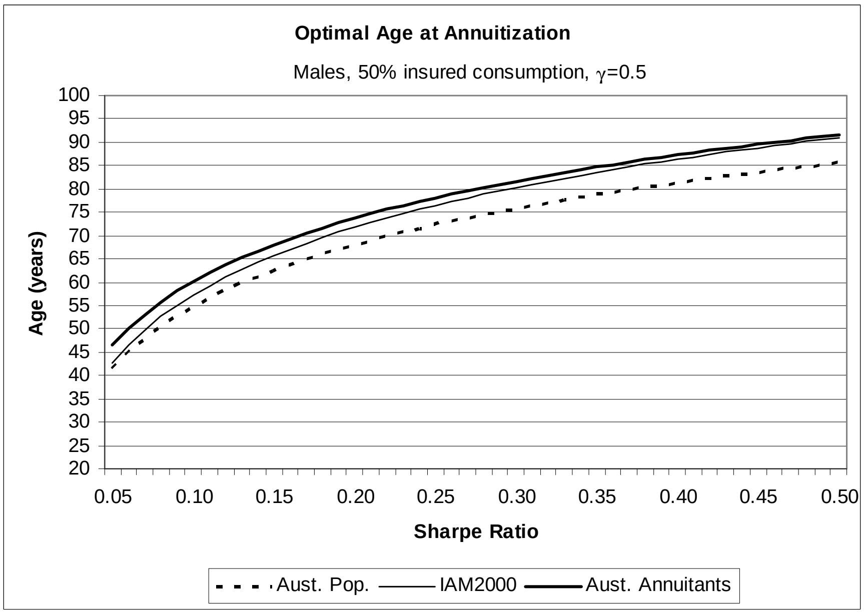


Figure 3

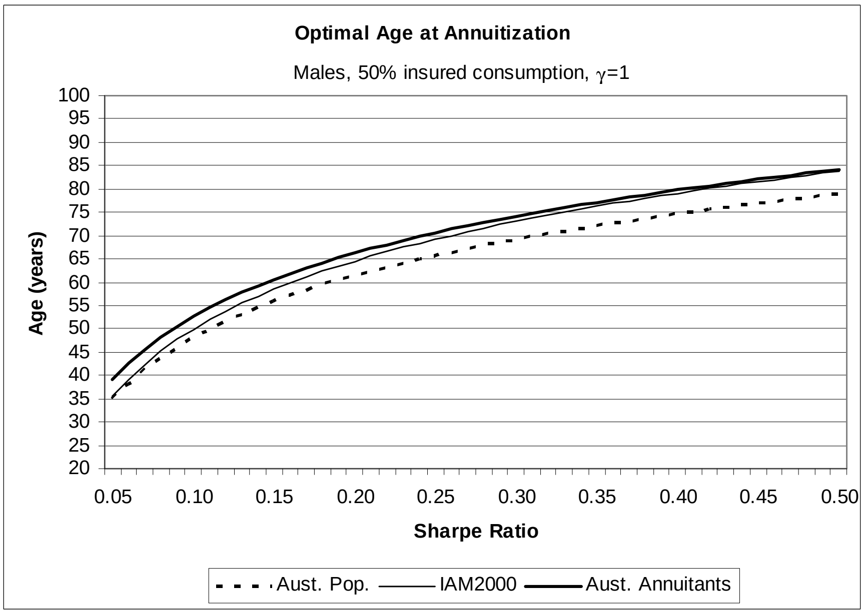
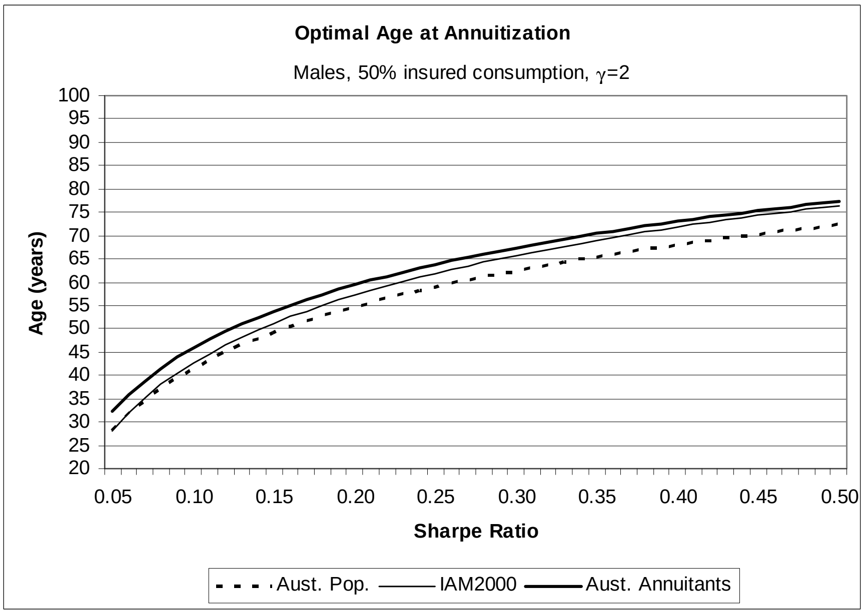


Figure 4



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.

Three 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 to 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. Thirdly, the presence of the stochastic wealth level in the condition defining optimal annuitization timing means that the real option to delay annuitization cannot be valued exactly from the initial period, since the timing depends on a random variable, and can be known with certainty only at the instant of annuitization.

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 and for individuals who have optimistic expectation of investment returns.

References

[1] Ameriks, J. and S.P. Zeldes (2001) 'How do portfolio shares vary with age?' Unpublished manuscript, TIAA-CREF Institute.

[2] Australian Government Actuary (1999) Australian Life Tables 1995-97, Commonwealth of Australia.

[3] Campbell, J.Y. (2002) 'Stock Returns for a New Century', TIAA-CREF Investment Forum, June, pp11-12.

[4] 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, vol. 107, no. 2 (April), pp.205-251.

[5] Carriere, J.F., (1992) 'Parametric models for life tables', Transactions of the Society of Actuaries 44, pp77-99.

[6] CCH (2002) Master Financial Planning Guide, 5th edn., CCH Australia: Sydney.

[7] Constantinides, George, (1990) 'Habit formation: a resolution of the equity premium puzzle.' Journal of Political Economy, vol. 98, no. 3, pp.519-543.

[8] 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.

[9] 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 Studies Discussion Paper no. 02/02, University of New South Wales, Sydney.

[10] Ibbotson, R.G. (2002), 'Building the Future from the Past', TIAA-CREF Investment Forum, June, pp10-12.

[11] Ingersoll, J.E. (1987) Theory of Financial Decision Making, Rowman & Littlefield, Savage.

[12] IFSA (2002) Retirement Income Streams Report, Investment and Financial Services Association, December.

[13] Jorion, P. and W. Goetzmann, 'Global stock markets in the twentieth century', Journal of Finance, vol. 54, June, pp.953-980.

[14] Kingston, G. (1989) 'Theoretical foundations of constant proportion portfolio insurance' Economics Letters, 29, pp.345-347.

[15] Knox, D.M. (2000) 'The Australian annuity market.' World Bank Working Paper no.2495, Development Research Group, World Bank.

[16] Leland, H.E. (1980) 'Who should buy portfolio insurance?', Journal of Finance, vol. 35, no. 2, pp.581-594.

[17] Liu, J. and E. Neiss (2002),'Endogenous retirement and portfolio choice', Working Paper, Anderson School of Business, UCLA.

[18] Merton, R.C. (1971) 'Optimum consumption and portfolio rules in a continuous time model', Journal of Economic Theory, vol. 3, pp.373-413.

[19] —— (1969) 'Lifetime portfolio selection under uncertainty: the continuous time case', Review of Economics and Statistics, 51, pp.239-246.

[20] Milevsky, M.A. and V.R. Young (2003) 'Annuitization and Asset Allocation', Management Science, forthcoming,.

[21] Milevsky, M.A. and V.R. Young (2002) 'Optimal asset allocation and the real option to delay annuitization: It's not now or never.', Discussion Paper PI-0211, The Pensions Institute, Birkbeck College, University of London.

[22] 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.

[23] Perold, A.F. and W.F. Sharpe (1988) 'Dynamic strategies for asset allocation', Financial Analysts Journal, January-February, pp.16-27.

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

[25] Rubinstein, M. (1976) 'The strong case for the Generalized Logarithmic Utility Model as the premier model of financial markets.' Journal of Finance 31, pp 551-71.

[26] Valdez, E.A., (2000) 'Developing a general law of mortality for Singapore', Singapore International Insurance and Actuarial Journal, June, pp1-17.

[27] Yaari, M. E., (1965) 'Uncertain lifetime, life insurance, and the theory of the consumer', Review of Economic Studies, 32, pp.137-150.

6. Appendix A: Derivation of Equation (14)

Specify the value function:

$$
\begin{aligned}
V(\tilde{w},t;T) &= A(t)^{\gamma}\frac{\tilde{w}}{1-\gamma}^{1-\gamma} && (6.1)\\
A(t)^{\gamma} &\equiv \left\{ \begin{array}{c} \left(\frac{\bar{a}^{b}_{x+T}}{\left(\bar{a}^{o}_{x+T}\right)^{1-\gamma}}\right)^{\frac{1}{\gamma}} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} \left({}_{T-t}p^{b}_{x+t}\right)^{\frac{1}{\gamma}} + \\ \int_{t}^{T} e^{-\frac{r-\delta(1-\gamma)}{\gamma}(s-t)} \left({}_{s-t}p^{b}_{x+t}\right)^{\frac{1}{\gamma}} ds \end{array} \right\}^{\gamma} && (6.2)\\
\hat{W}_{t} &\equiv \frac{\hat{C}}{r}\left(1-e^{r(t-T)}\right) + \hat{C}\bar{a}^{o}_{x+T}e^{r(t-T)} && (6.3)
\end{aligned}
$$

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

$$
\frac{\partial V}{\partial T} = \frac{\tilde{w}}{1-\gamma}^{1-\gamma}\gamma A(t)^{\gamma-1}\frac{\partial A(t)}{\partial T} + A(t)^{\gamma}\tilde{w}^{-\gamma}\frac{\partial \tilde{w}}{\partial T} \qquad (6.4)
$$

The first term in (6.4) is given by:

$$
\begin{aligned}
\frac{\tilde{w}}{1-\gamma}^{1-\gamma} \gamma A(t)^{\gamma-1} \frac{\partial A(t)}{\partial T} \quad = \quad & \frac{\tilde{w}}{1-\gamma}^{1-\gamma} \gamma A(t)^{\gamma-1} \\
& \text{x} \;\; e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)} \left({}_{T-t}p^{b}_{x+t} \right)^{\frac{1}{\gamma}} \frac{(1-\gamma)}{\gamma} \left(\frac{\bar{a}^{b}_{x+T}}{\bar{a}^{o}_{x+T}} \right)^{\frac{1-\gamma}{\gamma}} \\
& \text{x} \left\{ \begin{array}{l} \bar{a}^{b}_{x+T}(\delta - \lambda^{o}_{x+T} - r) - \frac{1}{(1-\gamma)} \\ + \frac{\bar{a}^{b}_{x+T}}{\bar{a}^{o}_{x+T}} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}^{b}_{x+T}}{\bar{a}^{o}_{x+T}} \right)^{-\frac{1-\gamma}{\gamma}} \end{array} \right\} \qquad (6.5)
\end{aligned}
$$

This expression is consistent with M-Y 2003 equation (16), which notes that without a consumption floor,

$$
\frac{\partial V}{\partial T} \propto \bar{a}^{b}_{x+T}(\delta - \lambda^{o}_{x+T} - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}^{b}_{x+T}}{\bar{a}^{o}_{x+T}} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}^{b}_{x+T}}{\bar{a}^{o}_{x+T}} \right)^{-\frac{1-\gamma}{\gamma}} \qquad (6.6)
$$

however, with a consumption floor the derivative has another additive term:

$$
A(t)^{\gamma} \tilde{w}^{-\gamma} \frac{\partial \tilde{w}}{\partial T}. \qquad (6.7)
$$

$$
\begin{aligned}
\frac{\partial \tilde{w}}{\partial T} &= \frac{\partial(-\hat{W}_t)}{\partial T} \\
&= -\left[\frac{\partial \frac{\hat{C}}{r}\left(1-e^{r(t-T)}\right)}{\partial T} + \frac{\partial \hat{C}\bar{a}^o_{x+T}e^{r(t-T)}}{\partial T}\right] \\
&= -\left[\hat{C}e^{r(t-T)} + \hat{C}\bar{a}^o_{x+T}. - r.e^{r(t-T)} + e^{r(t-T)}\hat{C}\frac{\partial}{\partial T}\bar{a}^o_{x+T}\right] \\
&= -\left[\hat{C}e^{r(t-T)}\left(1 - r\bar{a}^o_{x+T} + \left(\lambda^o_{x+T} + r\right)\bar{a}^o_{x+T} - 1\right)\right] \\
&= -\left[\hat{C}e^{r(t-T)}\lambda^o_{x+T}\bar{a}^o_{x+T}\right] \qquad (6.8)
\end{aligned}
$$

And

$$
\begin{aligned}
\frac{\partial V}{\partial T} &= \frac{\tilde{w}}{1-\gamma}^{1-\gamma}\gamma A(t)^{\gamma-1} \\
& \text{x} \;\; e^{-\frac{r-\delta(1-\gamma)}{\gamma}(T-t)}\left({}_{T-t}p^b_{x+t}\right)^{\frac{1}{\gamma}}.\frac{(1-\gamma)}{\gamma}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{\frac{1-\gamma}{\gamma}} \\
& \text{x} \left\{\bar{a}^b_{x+T}(\delta - \lambda^o_{x+T} - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}} + \frac{\gamma}{(1-\gamma)}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{-\frac{1-\gamma}{\gamma}}\right\} \\
& -A(t)^{\gamma}\tilde{w}^{-\gamma}\left[\hat{C}e^{r(t-T)}\lambda^o_{x+T}\bar{a}^o_{x+T}\right]. \qquad (6.9)
\end{aligned}
$$

By constraining $t = T$, note that

$$
A(T) = \left(\frac{\bar{a}^b_{x+T}}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}}\right)^{\frac{1}{\gamma}}
$$

hence

$$
\begin{aligned}
\left.\frac{\partial V}{\partial T}\right|_{t=T} &= \frac{\tilde{W}_T^{1-\gamma}}{1-\gamma}\gamma\left(\frac{\bar{a}^b_{x+T}}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}}\right)^{\frac{\gamma-1}{\gamma}}.\frac{(1-\gamma)}{\gamma}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{\frac{1-\gamma}{\gamma}} \\
&\quad \text{x}\left\{\bar{a}^b_{x+T}(\delta-\lambda^o_{x+T}-r)-\frac{1}{(1-\gamma)}+\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}+\frac{\gamma}{(1-\gamma)}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{-\frac{1-\gamma}{\gamma}}\right\} \\
&\quad -\frac{\bar{a}^b_{x+T}}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}}\tilde{W}_T^{-\gamma}\left[\hat{C}\lambda^o_{x+T}\bar{a}^o_{x+T}\right]. \\
&= \tilde{W}_T^{-\gamma}\left[\begin{array}{c} \frac{\tilde{W}_T}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}}\left\{\bar{a}^b_{x+T}(\delta-\lambda^o_{x+T}-r)-\frac{1}{(1-\gamma)}+\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}+\frac{\gamma}{(1-\gamma)}\left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{-\frac{1-\gamma}{\gamma}}\right\} \\ -\frac{\bar{a}^b_{x+T}}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}}\left[\hat{C}\lambda^o_{x+T}\bar{a}^o_{x+T}\right] \end{array}\right].
\end{aligned}
$$

And if $b = o$,

$$
\begin{aligned}
\left.\frac{\partial V}{\partial T}\right|_{t=T}^{b=o} &= \tilde{W}_T^{-\gamma}\left(\bar{a}_{x+T}\right)^{\gamma}\left[\tilde{W}_T(\delta-\lambda_{x+T}-r)-\hat{C}\lambda_{x+T}\bar{a}_{x+T}\right] && (6.10) \\
\left.\frac{\partial V}{\partial T}\right|_{t=T}^{b=o} &\propto \tilde{W}_T(\delta-\lambda_{x+T}-r)-\hat{C}\lambda_{x+T}\bar{a}_{x+T} && (6.11) \\
\left.\frac{\partial V}{\partial T}\right|_{t=T}^{b=o} &\propto (\delta-r)-\lambda_{x+T}\left(\frac{\hat{W}_T}{\tilde{W}_T}+1\right) && (6.12)
\end{aligned}
$$

7. Appendix B - 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 $\bar{a}^b_{x+T} < 2\bar{a}^o_{x+T}$. In the case where an individual views themselves as less likely to survive, this condition is always met because $\bar{a}^b_{x+T} < \bar{a}^o_{x+T}$, but the proof also holds for individuals who regard themselves as more likely to survive, as long as $\bar{a}^b_{x+T} < 2\bar{a}^o_{x+T}$. (Numerical examples give more general support for the result). In the case of HARA utility the divergence result still holds, conditional on all optimal annuitization dates being earlier than the equivalent $\hat{C} = 0$. The following adapts the M-Y proof to the HARA case.

The optimal moment of annuitization occurs when

$$\left.\frac{\partial V}{\partial T}\right|_{t=T} = \tilde{W}_T^{-\gamma} \left[\begin{array}{c} \frac{\tilde{W}_T}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}} \left\{ \bar{a}^b_{x+T}(\delta - \lambda^o_{x+T} - r) - \frac{1}{(1-\gamma)} + \frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}} + \frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{-\frac{1-\gamma}{\gamma}} \right\} \\ -\frac{\bar{a}^b_{x+T}}{\left(\bar{a}^o_{x+T}\right)^{1-\gamma}} \left[\hat{C}\lambda^o_{x+T}\bar{a}^o_{x+T}\right] \end{array} \right]$$

.

Setting this expression equal to zero gives:

$$0 = \bar{a}^b_{x+T}(\delta - \lambda^o_{x+T}(1 + \frac{\hat{W}_T}{\tilde{W}_T}) - r) + \left[\frac{\gamma}{(1-\gamma)} \left(\frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}}\right)^{-\frac{1-\gamma}{\gamma}} - \frac{1}{(1-\gamma)} + \frac{\bar{a}^b_{x+T}}{\bar{a}^o_{x+T}} \right] \quad (7.1)$$

.

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

$$\bar{a}^b_{x+T} = \bar{a}^o_{x+T} + \varepsilon \quad (7.2)$$

for small ε of either sign, and rewrite (7.1):

$$\begin{aligned} 0 &= (\bar{a}^o_{x+T}+\varepsilon)\left[\delta-\lambda^o_{x+T}(1+\frac{\hat{W}_T}{\tilde{W}_T})-r\right]+ \\ &\left[\frac{\gamma}{(1-\gamma)}\left(\frac{\bar{a}^o_{x+T}+\varepsilon}{\bar{a}^o_{x+T}}\right)^{-\frac{1-\gamma}{\gamma}}-\frac{1}{(1-\gamma)}+\frac{\bar{a}^o_{x+T}+\varepsilon}{\bar{a}^o_{x+T}}\right]. \end{aligned} \tag{7.3}$$

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

$$\begin{aligned} 0 &= (\bar{a}^o_{x+T}+\varepsilon)\left[\delta-\lambda^o_{x+T}(1+\frac{\hat{W}_T}{\tilde{W}_T})-r\right]-\frac{1}{2}\left(-\frac{1-\gamma}{\gamma}-1\right)\left(\frac{\varepsilon}{\bar{a}^o_{x+T}}\right)^2 \\ &-\frac{1}{6}\left(-\frac{1-\gamma}{\gamma}-1\right)\left(-\frac{1-\gamma}{\gamma}-2\right)\left(\frac{\varepsilon}{\bar{a}^o_{x+T}}\right)^3+..., \end{aligned} \tag{7.4}$$

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

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

$$\begin{aligned} 0 &= (\bar{a}^o_{x+T}+\varepsilon)\left[\delta-\lambda^o_{x+T}(1+\frac{\hat{W}_T}{\tilde{W}_T})-r\right]+\frac{(\varepsilon^*)^2}{2\gamma} \\ &= \left[\delta-\lambda^o_{x+T}(1+\frac{\hat{W}_T}{\tilde{W}_T})-r\right]+\frac{(\varepsilon^*)^2}{2\gamma}\frac{1}{\bar{a}^o_{x+T}\left(1+\frac{\varepsilon}{\bar{a}^o_{x+T}}\right)} \end{aligned} \tag{7.5}$$

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

$$\delta - r + \frac{(\varepsilon^*)^2}{2\gamma}\frac{1}{\bar{a}^o_{x+T}\left(1+\frac{\varepsilon}{\bar{a}^o_{x+T}}\right)} = \lambda^o_{x+T}(1+\frac{\hat{W}_T}{\tilde{W}_T})$$

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