Title of Academic Seminar:

Retirement Spending, Stochastic Mortality and Longevity Risk Aversion

Presented by: Moshe A. Milevsky
Schulich School of Business
York University, Toronto

(Joint work with Huaxiong Huang and Thomas Salisbury)

Enclosed are 2 papers that are related to the seminar presentation:

1.) Spending Retirement on Planet Vulcan: The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates (last revised: August 2010)

2.) Calibrating Yaari to the 21st Century, Part I: Consumption Under a Stochastic Force of Mortality (last revised: November 2010)
SPENDING RETIREMENT ON PLANET VULCAN:
The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates

Professor Moshe A. Milevsky, Ph.D.
milevsky@yorku.ca
Schulich School of Business, York University
& The IFID Centre
Toronto, Ontario, CANADA, M3J 1P3

&

Professor Huaxiong Huang, Ph.D.
hhuang@yorku.ca
Department of Mathematics and Statistics, York University
Toronto, Ontario, CANADA, M3J 1P3

Final Version: 23 August 2010
SPENDING RETIREMENT ON PLANET VULCAN:  
*The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates*

Abstract: [50 words only]

Recommendations offered by the media and financial planners on retirement spending rates deviate considerably from utility-maximization models. For example, the widely quoted 4% target-spending rule is incompatible with lifecycle consumption smoothing. Rather, wealth managers should advocate dynamic spending in proportion to survival probabilities, adjusted (up) for exogenous pension income and (down) for longevity risk aversion. We explain exactly why and how.
Online Summary [450 words only]

In this article we illustrate how to derive, analyze and explain the optimal retirement spending policy for a utility-maximizing consumer facing (only) a stochastic lifetime. We deliberately and consciously ignore financial market risk by assuming that all investment assets are allocated to risk-free (e.g. TIPS) bonds. This simplifying assumption is made in order to help focus reader’s attention on the role of “longevity risk aversion” on optimal consumption and spending rates during a retirement period of stochastic length.

Indeed, the impact of financial risk aversion on optimal asset allocation has been the subject of many articles and intuitively well understood. Yet, the impact of longevity risk aversion on retirement spending rates has not received as much attention, nor are most practitioners even familiar with this concept. Over 75 million baby-boomers are (still) hoping to retire one day – each with their own stochastic remaining lifespan -- and are likely to be demanding advice from their wealth managers on this exact issue.

And, while neither our framework nor mathematical solution is original – it can actually be traced back almost 100 years – we believe the insights from a normative lifecycle model are worth emphasizing in the current environment that has grown jaded of economic models and their prescriptions. In particular, we contrast the optimal (a.k.a. utility maximizing) retirement spending policy with popular recommendations offered by the investment media and financial planners.

Our main point of contention is that counseling retirees to set initial spending from investable wealth at a constant inflation-adjusted rate – for example the widely popularized 4% rule -- is consistent with lifecycle consumption smoothing only under a very limited set of implausible preference parameters. There simply is no universally optimal or safe retirement spending rate.

Rather, the optimal forward-looking behavior in the face of personal longevity risk is to consume in proportion to survival probabilities – adjusted (upwards) for pension income and (downwards) for longevity risk aversion -- as opposed to blindly withdrawing constant income for life. This framework also allows one to illustrate the impact (and benefit) of pension income annuities on the optimal plan.

We believe that 21st century wealth managers who have grown accustomed to framing all their discussions with clients around the prism of risk and return, should advocate dynamic spending policies in this manner. Thus, our intent with this paper isn’t to dismiss or belittle widely-used rules of thumb, but rather to create a common language and help improve the dialogue between financial economists and the financial planning community. The stakes are simply too high to allow yet another naive rule of thumb to take-hold in these complex and uncertain environments.
INTRODUCTION:

In this article we illustrate how to derive, analyze and explain the optimal retirement spending policy for a utility-maximizing consumer facing (only) a stochastic lifetime. We deliberately and consciously ignore financial market risk by assuming that all investment assets are allocated to risk-free (e.g. TIPS) bonds. This simplifying assumption is made in order to focus reader’s attention on the role of “longevity risk aversion” on optimal consumption and spending rates during a retirement period of stochastic length.

Loosely speaking, by longevity risk aversion we mean to imply that different people might have differing attitudes towards the “fear” of living longer than anticipated and possibly depleting their financial resources. Some might respond to this risk by spending less early-on in retirement, while others might be willing to take their chances and enjoy a higher standard of living while they are most likely to be alive.

Indeed, the impact of financial risk aversion on optimal asset allocation has been the subject of many articles and is intuitively well understood by practitioners. Investors who are more worried and concerned about losing money (a.k.a risk averse) invest more conservatively, sacrifice the potential upside which in effect leads to a reduced lifetime standard of living. In contrast, the impact of longevity risk aversion on retirement spending behavior has not received as much attention, nor are most financial practitioners even familiar with this concept.

And, while neither our framework nor mathematical solution is original – as we will show in the next section, it can actually be traced back almost 80 years – we believe the insights from a normative lifecycle model are worth emphasizing in the current environment that has grown jaded of economic models and their prescriptions. Our pedagogical objective is contrast the optimal (a.k.a. utility maximizing) retirement spending policy with popular recommendations offered by the investment media and financial planners.

Our main point of contention is that counseling retirees to set initial spending from investable wealth at a constant inflation-adjusted rate – for example the widely popularized 4% rule -- is consistent with lifecycle consumption smoothing only under a very limited set of implausible preference parameters. There simply is no universally optimal or safe retirement spending rate.

Rather, the optimal forward-looking behavior in the face of personal longevity risk is to consume in proportion to survival probabilities – adjusted (upwards) for pension income and (downwards) for
longevity risk aversion -- as opposed to blindly withdrawing constant income for life. More on this later; first we offer some historical perspective on the lifecycle spending problem.

**History of the Problem:**

"...The first problem I propose to tackle is this: how much of its income should a nation save?..." With these words the 24-year-old Cambridge University economist Frank R. Ramsey began a celebrated paper published in the *Economic Journal* two years before his tragic death in 1930. The so-called Ramsey (1928) model and the resultant Keynes-Ramsey rule, implicitly subsumed by thousands of economists in the last 80 years -- including Fisher (1930), Modigliani (1984, 1956), Phelps (1962) and eventually Yaari (1965) -- represents the foundation for lifecycle utility optimization. It is also the workhorse supporting the original asset allocation models of Samuelson (1969) and Merton (1971).

In its basic form the normative lifecycle model (LCM) assumes a rational individual who seeks to maximize the discounted additive utility of consumption over their entire life. Despite its macro-economic origins the Ramsey (1928) model has been extended by thousands of economists. Indeed, ask a first-year graduate student in economics how a consumer should be “spending” over some deterministic time horizon T, and most likely they will respond with a Ramsey-type model spreading human capital and financial capital (i.e. total wealth) between time zero and the terminal time T.

The finance-oriented literature has advanced since 1928, and now falls under the title of “portfolio choice” or extensions of the Merton model. We count over 50 scholarly articles on this topic published in the top scholarly journals in finance over the last decade alone. *Unfortunately, much of the financial planning community has ignored these dynamic optimization models and nowhere is this more evident than in the “retirement income planning” world.*

Lamentably, the financial crisis coupled with general skepticism towards economic models has moved the practice of personal finance even farther away from a dynamic optimization approach. In fact, many popular and widely advocated strategies are at odds with the prescriptions of financial economics. See the articles by Bodie and Treaussard (2007), Kotlikoff (2008) as well as the monograph edited by Bodie, McLeavey and Siegel (2008) or the recent book by Ayres and Nalebuff (2010) for examples of how economists “think about” problems in personal finance and how it differs from conventional wisdom.
Along these same lines, we attempt to reconcile the gap between the advice of the financial planning community regarding retirement spending policies vis-à-vis the “advice” of financial economists using a rational utility-maximizing model of consumer choice\(^1\).

In particular, we focus exclusively on the impact of lifetime uncertainty – longevity risk – on the optimal consumption and retirement spending policy. To isolate the impact of longevity risk on optimal portfolio withdrawal rates in retirement, we have placed our deliberations on Planet Vulcan – where investment returns are known and unvarying; the inhabitants are rational, utility-maximizing consumption smoothers, and only lifespans are random.

**LITERATURE ON RETIREMENT SPENDING RATES:**

Within the community of retirement income planners, an often-referenced study is the work by Bengen (1994) in which he used historical (Ibbotson Associates) equity and bond returns to search for the highest allowable spending rate that would sustain a portfolio for 30 years of retirement. Using a 50/50 equity/bond mix, Bengen settled on a rate between 4% and 5%. In fact, this is known as the Bengen or 4% Rule in the retirement income planning community, and has caught-on like a wildfire. For those unfamiliar with this rule, is simply says that for every $100 in the retirement nest egg, one should withdraw $4 adjusted for inflation each year – forever, or at least until the portfolio runs dry or life runs out, whichever comes first.

Indeed, it is hard to over-estimate the influence of the Bengen (1994) article and its embedded “rule” on the contemporary practice of retirement income planning. Other studies in the same vein include an article in the *Journal of the American Association of Individual Investors* (AAII), by Cooley, Hubbard, and Walz (1998) often referred to as the Trinity Study. These and related studies have been quoted and cited thousands of times – by the popular media -- in the last two decades (Money Magazine, USA Today, Wall Street Journal)\(^2\). The 4% spending rule now seems destined for the same immortality enjoyed by other (overly simplistic) rules of thumb such as buy-term-and-invest-the-difference or dollar-cost

\(^1\) Hence our use of “Planet Vulcan” in our title, inspired by Richard Thaler and Cass Sunstein who distinguish “humans” from perfectly rational “econs,” much like Spock in TV’s Star Trek, who originates from the Planet Vulcan.

\(^2\) Money Magazine, August 16, 2007 article by Walter Updegrave, or CBS MarketWatch, July 29, 2009 article by Janet Kidd Stewart.
averaging. And, while numerous authors have extended, refined and re-calibrated these spending rules, their spirit remains intact across all versions.3

Yet this “start by spending x%” strategy has no basis in economic theory. We are not the first to point this out. For example, a series of articles by Sharpe, Scott and Watson (2007, 2009) raised similar concerns and alluded to the need for a life-cycle approach, but without solving or calibrating such a model. Our goal here is to actually illustrate the solution to the life-cycle problem and demonstrate how longevity risk aversion — in contrast to financial risk aversion, so familiar to financial analysts — impacts retirement spending rates.

Recently published articles have teased-out the implications of mortality and longevity risk on portfolio choice and asset allocation, for example the paper by Chen, et al. (2006) in the Financial Analysts Journal (FAJ). Likewise, the paper by Milevsky and Robinson (2005) argued that retirement spending rates should be lower because the embedded equity risk premium (ERP) assumption is too high. In contrast to these approaches, our paper uses an economic life-cycle approach to retirement income planning.

With the context and motivation out of the way, the remainder of this paper is organized as follows. The lifecycle model itself is developed in the appendix. The section entitled “Numerical Examples” which follows, displays our results. The subsequent section titled “Summary and Conclusion” provides a qualitative summary of our main insights and some big-picture takeaways. All references are at the end of the paper.

NUMERICAL EXAMPLES AND CASES

The model we use is fully described in the appendix so that readers can select their own parameter values and derive optimal spending rates under any assumptions. With our equations given, this can actually be done quite easily in Excel. We have selected one (plausible) set of values to illustrate the main qualitative insights which are rather insensitive to actual values.

Recall that we are spending our retirement on Planet Vulcan where only life-spans are random. Our approach forces us to specify a real (inflation-adjusted) investment return. So, after carefully examining the real yield from U.S. TIPS over the last ten years, based on data from the Federal Reserve. The

3 In fact, we (the authors) are partially guilty of helping proliferate this approach by deriving analytic expressions for the portfolio ruin probability assuming a constant consumption rate.
maximum real yield over the period was 3.15% for the 10-year bond, and 4.24% for the 5-year bond. The average yield was 1.95% and 1.50% respectively. The longer maturity TIPS exhibited higher yields, but obviously entail some duration risk. So, after much deliberation we decided to use a real interest rate assumption of 2.5% for most of the examples, even though current (Fall 2010) TIPS rates are substantially lower. Our values are consistent with a view expressed by Arnott (2004) regarding the future of the equity risk premium (ERP).

As far as longevity risk is concerned, we assumed the retiree’s remaining lifetime obeys a (unisex) biological law of mortality under which the hazard rate increases exponentially over time. This is known as the Gompertz assumption in the actuarial literature and we calibrated this model to common pension (RP2000) tables. The mortality law is described fully in the appendix.

What this means is that in most of our numerical examples we assume an 86.6% probability that a 65-year-old will survive to the age of 75, a 57.3% probability of reaching 85, a 36.9% probability of reaching 90, a 17.6% probability of reaching age 95 and a 5% probability of reaching 100. Note again that we do not plan for a life expectancy or an ad hoc 30 year retirement. Rather, we account for the entire term structure of mortality.

Because the main objective of this paper is to focus attention on the impact of risk aversion on the optimal portfolio withdrawal rates (PWR), we display results for a range of values. We show results for a retiree with a very low ($\gamma=1$) and a relatively high ($\gamma=8$) coefficient of relative risk aversion (CRRA).

To better understand mortality-risk aversion, we offer the following analogy to classical asset allocation models. An investor with a CRRA value of ($\gamma=4$) would invest 40% of his or her assets in an equity portfolio and 60% in a bond portfolio, assuming the equity risk premium is 5% and volatility is 18%. Our model does not have a risky asset nor does it require an ERP, but the idea is that the CRRA can be mapped onto more easily understood risk attitudes. Along the same lines, the very low (a.k.a. Bernoulli) risk aversion value of ($\gamma=1$) would lead to an equity allocation of 150%, and a high risk aversion value of ($\gamma=8$) implies an equity allocation of 20%.

Finally, to complete the parameter values required for our model, we assume the subjective discount rate ($\rho$) which is a proxy for personal impatience, is equal to the risk-free rate (mostly 2.5% in our numerical examples). To those familiar with the basic lifecycle model without lifetime uncertainty, this
implies that the optimal consumption rates would be constant over time in the absence of longevity risk considerations.

In the language of economics, when the Subjective Discount Rate (SDR) in a lifecycle model is set equal to the constant and risk-free interest rate, then a rational consumer would spend their total (human plus financial) capital evenly and in equal amounts over time. In other words, in a model with no horizon uncertainty, consumption rates and spending amounts are in fact constant, regardless of the consumer’s Elasticity of Inter-temporal Substitution (EIS).4

The question of interest is: What happens when lifetimes are stochastic?

<table>
<thead>
<tr>
<th>Initial Portfolio (Nest Egg)</th>
<th>Real Interest Rate = 0.5%</th>
<th>Real Interest Rate = 1.5%</th>
<th>Real Interest Rate = 2.5%</th>
<th>Real Interest Rate = 3.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retire at Age 65</td>
<td>$3.330</td>
<td>$3.941</td>
<td>$4.605</td>
<td>$5.318</td>
</tr>
<tr>
<td>5 Years Later</td>
<td>$3.286</td>
<td>$3.888</td>
<td>$4.544</td>
<td>$5.247</td>
</tr>
<tr>
<td>10 Years Later</td>
<td>$3.212</td>
<td>$3.801</td>
<td>$4.442</td>
<td>$5.130</td>
</tr>
<tr>
<td>20 Years Later</td>
<td>$2.898</td>
<td>$3.429</td>
<td>$4.007</td>
<td>$4.627</td>
</tr>
<tr>
<td>30 Years Later</td>
<td>$2.156</td>
<td>$2.552</td>
<td>$2.982</td>
<td>$3.444</td>
</tr>
</tbody>
</table>

Note: 5% survival probability to age 100. Gompertz Mortality with Parameters (m=89.335, b=9.5). No pension income assumed. All consumption takes place from investment portfolio.

We are now ready for some results. Assume a 65-year-old with a (standardized) $100 nest egg. Initially we allow for no pension annuity income and therefore all consumption must be sourced to the investment portfolio which is earning a deterministic interest rate. On Planet Vulcan financial wealth must be depleted at the very end of the lifecycle (age 120) and there are no bequest motives. So, according to equation (#5) in the technical appendix, the optimal consumption rate at retirement age 65 is $4.605 when the risk aversion parameter is set to ($\gamma=4$), displayed in Table #1, and the optimal consumption rate is $4.121$ when the risk aversion parameter is set to ($\gamma=8$).

4 For those interested in more detailed information about possible parameter estimates for the EIS and how this impacts consumption in deterministic lifecycle models for which the SDR is not equal to the interest rate, we refer to the paper by Hanna, Fan and Chang (1995).
Notice that these are within the range of numbers quoted by the popular press for optimal portfolio withdrawal (spending) rates. Thus, at first glance this seems to suggest that simple 4% rules of thumb are consistent with a lifecycle model. Unfortunately, the euphoria is short-lived. It is only in the first year of withdrawals, age 65, in which the numbers (might) coincide. As the retiree ages they rationally consume less each year – in proportion to their survival probability adjusted for risk aversion. For example, in our baseline intermediate ($\gamma=4$) level of risk aversion, the optimal consumption rate drops from $\$4.605$ at age 65, to $\$4.544$ at age 70, then $\$4.442$ at age 75, then (not displayed in table) $\$3.591$ at age 90 and $\$2.177$ at age 100, assuming the retiree is still alive. All of these values come from equation (5) in the appendix.

Note how a lower real interest rate, for example 0.5% in Table #1, leads to a reduced optimal retirement consumption/spending rate. Indeed, in the current (Fall 2010) yield curve and TIPS environment our model has yet an important message for baby boomers: Your parent’s retirement plan might not be sustainable anymore.

Either way, the first insight in our model is that a fully rational plan is to actually spend less as you progress through retirement. The lifecycle optimizer (i.e. “consumption smoother” on Planet Vulcan) spends more at earlier ages and reduces spending as they age, even if their subjective discount rate (SDR) is equal to the real interest rate in the economy.

Intuitively they deal with longevity risk by setting aside a financial reserve AND by planning to reduce consumption -- if that risk materializes -- in proportion to the survival probability, linked to their risk aversion. All of this is in the absence of any pension annuity income.

To quote Irving Fisher (1930) in his Theory of Interest (page 85): “...The shortness of life thus tends powerfully to increase the degree of impatience or rate of time preference beyond what it would otherwise be...” and (page 90) “Everyone at some time in his life doubtless changes his degree of impatience for income...When he gets a little older,...he expects to die and he thinks: instead of piling up for the remote future, why shouldn’t I enjoy myself during the few years that remain?”

Including Pension Annuities

We now employ the same model to examine what happens when the retiree has access to a Defined Benefit (DB) pension income annuity, which provides a guaranteed lifetime cash-flow. The maximum amount of Social Security (S.S.) in the U.S., which is the ultimate real pension annuity, is approximately
$25,000 per individual. We examine the behavior of a retiree with 100, 50 and 20 times this amount in their nest egg: we consider (i.) $2,500,000, (ii.) $1,250,000 and (iii.) $500,000 in investable retirement assets.

Alternatively, one can interpret Table #2 as displaying the optimal policy for 4 different retirees or varying degrees of risk tolerance and aversion, each with $1,000,000 in investable retirement assets. The first has no ($\pi_0=$0) pension, the second has a pension of $10,000 per year ($\pi_0=$1), the third has a pension of $20,000 per year ($\pi_0=$2) and the fourth has a pension of $50,000 ($\pi_0=$5).

<table>
<thead>
<tr>
<th>Pension Income</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 2$</th>
<th>$\gamma = 4$</th>
<th>$\gamma = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_0 = $0$</td>
<td>6.330%</td>
<td>5.301%</td>
<td>4.605%</td>
<td>4.121%</td>
</tr>
<tr>
<td>$\pi_0 = $1$</td>
<td>6.798%</td>
<td>5.653%</td>
<td>4.873%</td>
<td>4.324%</td>
</tr>
<tr>
<td>$\pi_0 = $2$</td>
<td>7.162%</td>
<td>5.924%</td>
<td>5.078%</td>
<td>4.480%</td>
</tr>
<tr>
<td>$\pi_0 = $5$</td>
<td>8.015%</td>
<td>6.553%</td>
<td>5.551%</td>
<td>4.839%</td>
</tr>
</tbody>
</table>

Note: 5% survival probability to age 100. Equivalent to Gompertz (m=89.335, b=9.5). Interest Rate = 2.5%

Table #2 displays the net portfolio withdrawal rates as a function of the risk aversion values and pre-existing pension income. By net portfolio withdrawal rates (PWR) we mean the optimal amount withdrawn from the investment portfolio.

Thus, for example, when the ($\gamma = 4$) -- e.g. medium risk aversion -- retiree has $1,000,000 in investable assets and is entitled to a real lifetime pension of $50,000, which using our language is a scaled nest egg of $100 and a pension ($\pi_0=$5), the optimal total consumption rate is $10.551 in the first year. Of this sum, $5.00 obviously comes from the pension and $5.551 is withdrawn from the portfolio. Thus, the net portfolio withdrawal rate (PWR) is 5.551%.

In contrast, if the retiree has the same $1,000,000 in assets but only entitled to $10,000 in lifetime pension income, then the optimal total consumption rate is $5.873 per $100 of assets at age 65, of which $1.00 comes from the pension and $4.873 is withdrawn from the portfolio. Hence, the PWR is 4.873%. All these numbers come directly from equation (#5) in the technical appendix.
So, here for all intents and purposes is the main point of our paper in one summary sentence. The optimal net portfolio withdrawal rate (PWR) depends on longevity risk aversion and the level of pre-existing pension income. The larger the amount of the pre-existing pension income, the greater is the optimal consumption rate and the greater is the PWR.

Basically, the pension acts as a buffer and allows the retiree to consume more from discretionary wealth. Even at high levels of longevity risk aversion, the risk of living a longer lifespan doesn’t “worry” the retiree as much, since they have pension income to fall back upon should that chance (i.e. a long life) materialize. We believe this insight is absent from most of the popular media discussion (and practitioner implementation) of optimal spending rates. If a potential client has substantial pension income from a Defined Benefit (DB) pension, or Social Security, they can afford to withdraw more – percentagewise – compared to their neighbor who is relying entirely on their investment portfolio to finance their retirement income needs.

Table #2 confirms a number of other important results. Notice how the optimal portfolio withdrawal rate – for a range of pension income and risk aversion levels – is in fact between 8% and 4% but only when the inflation-adjusted interest rate is assumed to be a rather generous 2.5%. Adding another 100 basis points to the investment return assumption adds somewhere between 60 to 80 basis points to the initial PWR. However, reducing interest assumptions will have the opposite effects. Readers can input their own assumptions into equation #5 in the appendix to obtain suitable consumption/spending rates.

The impact of longevity risk aversion can alternatively be described as follows. If the remaining future lifetime has a modal value of \( m = 89.335 \) and a dispersion (volatility) value of \( b = 9.5 \), then a longevity-risk averse consumer behaves (consumes) as if the modal value was: \( m^* = m + b \ln \gamma \), but with the same dispersion parameter \( b \).

Longevity risk aversion manifests itself by (essentially) assuming you will live longer than the biological/medical estimate. It is only extremely risk tolerant retirees \( (\gamma = 1) \) who behave as if their modal lifespan is the true (biological) modal value. Note that this is not risk-neutrality, which would ignore longevity risk all together.

The closest analogy to these risk-adjusted mortality rates, within the asset allocation literature is the concept of risk-adjusted investment returns. A risk-averse investor observes a 10% expected portfolio return and adjusts it downward based on the volatility of the return and their risk aversion. If the (subjectively) adjusted investment return is under the risk-free rate, the investor shuns the risky asset.
Of course this analogy isn’t quite correct since the retiree can’t shun longevity risk, but the spirit is the same. The longevity they see isn’t the longevity they feel.

An important takeaway here (again) is the impact of pension annuities on retirement consumption. While the point of this paper is not to advocate for pension annuities or examine the marker for longevity protection – that is well achieved in the recent book by Sheshink (2008) as well as the excellent collection of articles by Brown, Mitchell, Poterba and Warshawsky (2001) – here is yet another way to use equations (#5, #6) in the appendix.

Table #3 displays the optimal consumption rate at various ages assuming a fixed percentage of the retirement nest egg is used to purchase a pension annuity, a.k.a. pensionized. The cost of each lifetime dollar of income is displayed in equation (#2) in the appendix, which is the expression for the pension annuity factor. So, if 30% of $100 is “pensionized” the corresponding value of \( F_0 = $70 \) and resulting pension annuity income is $30/\mu_{65}^{50}(0.025,89.335,9.5) = $1.899.

We note that the pricing of pension (income) annuities by private sector insurance companies will usually involve mortality rates that differ from population rates, due to anti-selection concerns. This could be easily incorporated by using different mortality parameters, but for now we keep things simple to illustrate the impact of lifetime income on optimal total spending rates.

Results are displayed at retirement age 65 and planned consumption 15 years later (assuming the individual is still alive) at age 80. We illustrate a variety of different scenarios in which 0%, 20%, 40%, 60% and 100% of initial wealth is pensionized. Once again by “pensionization” we mean the purchase of a non-reversible pension annuity – priced by equation (#2) – based on the going market rate.\(^5\)

<table>
<thead>
<tr>
<th>TABLE #3: HOW DOES “PENSIONIZATION” IMPACT RETIREMENT CONSUMPTION?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of $100 Pensionized</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>0%</td>
</tr>
</tbody>
</table>

\(^5\) This is quite different from the Yaari (1965) tontine annuity in which mortality credits are paid-out instantaneously by adding the mortality hazard rate \( \lambda_t \) to the investment return \( r \). This is why we use the phrase “pensionization” to distinguish from the economist’s use of the term annuitization. The latter assumes a pool in which survivors inherit the assets of the deceased, while the former requires an insurance company or pension fund to guarantee the lifetime payments. See the companion paper by Huang and Milevsky (2010) for a discussion of the distinction between the two and its impact on optimal retirement planning.
Table #3 displays total dollar consumption rates, including the corresponding pension annuity income. These are not (only) the portfolio withdrawal rates which were displayed in percentages in Table #2. So, for example, if the (medium risk aversion) retiree allocates $20 (from the $100 available) to purchase a pension annuity that pays $1.261 for life, then optimal consumption will be $1.261 + $3.997 = $5.263 at age 65. Note that the $3.997 withdrawn from the remaining portfolio of $80 is equivalent to an initial portfolio withdrawal rate of 4.997%.

In contrast, the retiree with a high degree of longevity risk aversion (\( \gamma = 8 \)) will receive the same $1.261 from the $20 that has been “pensionized” but will only (optimal) spend $3.535 from the portfolio (a withdrawal rate of 4.419%), for a total consumption rate of $4.801 at age 65.

And, if the entire nest egg is pensionized at 65, leading to $6.330 of lifetime income, then the consumption rate is constant for life – and independent of risk aversion -- as there is no financial capital from which to draw down any income. This is yet another way to illustrate the benefit of converting financial wealth into a pension income flow. The $6.330 of annual consumption is the largest of all consumption plans. This, by the way, is why most financial economists are strong advocates of “pensionizing” (or at least annuitizing) a portion of one’s retirement nest egg.
Visualizing the Results

Figure #1 displays the optimal consumption path from retirement until the maximum length of life as a function of the retiree's level of longevity risk aversion (remember: \(\gamma\) in our model). This figure gives yet another perspective on the rational approach and attitude to longevity risk management. Figure #1 uses equation (6) to trace-out the entire consumption path from retirement at age 65 until age 100.

Notice how the optimal consumption rate declines with age and in relation to the retiree's attitude towards longevity risk as measured by their Coefficient of Relative Risk Aversion (CRRA).

The figure plots four cases corresponding to differing levels of the CRRA. Notice how the consumption rate eventually hits $5, which is the pension income flow. So, for example, the CRRA = 2 (i.e. very low aversion to longevity risk) consumer will start retirement by withdrawing 6.55% from their nest egg plus their pension income of $5. The withdrawals from the portfolio will continue until they (rationally)
exhaust their wealth at age 95. From what is coined the Wealth Depletion Date (WDT) onwards, all they consume is their pension\textsuperscript{6}.

Figure #2 displays the corresponding trajectory for financial capital. At all levels of longevity risk aversion the curve begins at $100 and then declines. The rate of decline is higher and faster for lower levels of longevity risk aversion -- because the individual is “not afraid” of living to an advanced age. They will deplete their wealth after 24.6 years (which is age 90) after which they live on their pension ($5).

![Figure 2: Financial Capital: $5 Pension Income and Investment Rate= 2.5%](image)

In contrast, the retiree with longevity risk aversion CRRA = 8 doesn’t (plan to) deplete wealth until age 105, and draws down wealth at a much slower rate. When there is no pension annuity income at all, the wealth depletion time is exactly at the end of terminal horizon, which is the last possible age on the mortality table. In other words, wealth is never completely exhausted. This can also be seen from equation (\#8) where the only way to get zero (on the right hand side) is when the survival probability is zero, which can only happen when $\tau$ is equal to the maximum length of life.

\textsuperscript{6} For those readers who are interested, the consumption function is concave until the WDT, at which point it is non-differentiable and forced equal to the pension annuity income.
Reacting to Financial Shocks

Our methodology also allows us to examine the optimal reaction to financial shocks over the retirement horizon. Take someone who experiences a 30% loss in their investment portfolio and wants to rationally reduce spending to account for the depleted nest egg. The rule of thumb which suggests that a retiree spend 4% to 5%, says nothing about how to update this rule in response to a shock to wealth.

The rational reaction to a financial shock at time $s$, which results in a new (reduced) portfolio value, would be to follow these steps:

1. Recalibrate the model from time zero, but with the shocked level of wealth and compute the new wealth depletion time from equation (#8).
2. Use equation (#7) to compute the new level of initial consumption, which will be different from (the old) consumption level because of the financial shock.
3. Continue retirement consumption from time $s$ onward, based on equation (#5).

To understand how this would work in practice, let’s begin with a (CRRA=4) retiree who has $100 in investable assets and is entitled to $2 of lifetime pension income. Under an $r = 2.5\%$ real interest rate, the optimal policy is to consume a total of $7.078$ at age 65 ($2$ from the pension and $5.078$ from the portfolio) and adjust withdrawals downwards over time in proportion to the survival probability to the power of the risk-aversion coefficient. The wealth depletion time is at age 105. This has been explained previously.

Under this dynamic policy, the expectation is that at age 70 the financial capital trajectory will be $86.668$ and total consumption will be $6.984$, if the retiree follows the optimal consumption path for the next five years.

Now assume the retiree survives five years and experiences a financial shock which reduces the portfolio value from the expected $86.668$ to $60$ at age 70, which is $31\%$ less than planned. In this case the optimal plan is to reduce consumption to $5.583$, which is obtained by solving the problem from the beginning, but with a starting age of 70. This is a reduction of approximately $20\%$ compared to the original plan.

Of course, there is somewhat of an apples to oranges comparison here since (i.) a shock is not really allowed in our model, and (ii.) the time zero consumption plan is based on a conditional probability of survival that might change based on realized health status. This is the problem of stochastic (versus
deterministic) investment returns and mortality hazard rates, which obviously takes us far beyond the simple agenda for this paper.

In sum, all we want to say here is as follows: a rational response to an x% drop in one’s retirement portfolio is not to reduce consumption and spending by the same x%. This is a direct result of consumption smoothing in the lifecycle model.

![Economic Tradeoffs at Retirement](image)

**SUMMARY AND CONCLUSION:**

To a financial economist, the optimal retirement consumption rate, asset allocation (investments) and product allocation (insurance) is a complicated function of mortality expectations, economic forecasts and the tradeoff between the preference for retirement sustainability versus the desire to leave a financial legacy (bequest motive). It is not an easy problem to solve even under some very simplifying assumptions. But, the qualitative tradeoff can be illustrated by Figure #3

A retiree can afford to spend more if they are willing to leave a smaller financial legacy and risk early depletion times. They should spend less if they desire a larger legacy and greater sustainability. Optimization of investments and insurance products takes place on this retirement income frontier. Ergo, a simple rule that advises all retirees to spend x% of their nest egg adjusted up or down in some *ad hoc* manner is akin to the broken clock which tells time correctly only twice a day.
Naturally, we are not the first authors and certainly will not be the last to criticize the “spend x%” approach to retirement income planning. For example, the Professor William Sharpe and co-authors (2009) wrote:

“...The 4% rule and its variants finance a constant, non-volatile spending plan using a risky, volatile investment strategy. Two of the rule’s inefficiencies—the price paid for funding its unspent surpluses and the overpayments for its spending distribution—apply to all retirees, independent of their preferences....”

We obviously concur, but our focus in this paper is to illustrate what a lifecycle model actually says about optimal consumption rates. Our intention was to contrast ad hoc recommendations with “advice” that a financial economist would give to a utility-maximizing consumer, and see if there is any overlap and by how much, exactly, it differs.

In particular, we shine light on the aversion to longevity risk – the uncertainty of human lifespan – and to examine how this impacts optimal spending rates.7

Computationally we solved (in the appendix) an analytic life-cycle model (LCM) that was calibrated to actuarial mortality rates. Our model can easily be used by anyone with access to an Excel spreadsheet, and our main insights are as follows:

1. The optimal initial portfolio withdrawal rate (PWR) which the “planning literature” advocates should be an exogenous percentage of one’s retirement nest egg, actually depends quite critically on both the consumer’s risk aversion – where risk is longevity and not just financial markets – as well as any pre-existing pension annuity income. For example, if the portfolio’s real investment return is 2.5% per annum, then for individuals who are highly risk-averse the optimal initial PWR can be as low as 3%, and for individuals who are less risk-averse it can be as high as 7%. The same applies to the existence of pension annuity income. The greater the amount of pre-existing pensions the larger the initial PWR, all else being equal. Of course, if one assumes a healthier retiree and/or lower inflation-adjusted returns the optimal initial PWR is lower.

2. The optimal consumption rate (denoted by $c_t^*$), which is the total amount of money consumed by the retiree in any given year including all pension income, is a declining function of age. In

---

7 A possible (tongue in cheek) rule of thumb that could substitute for the static 4% algorithm is to counsel retirees to pick any initial spending rate between 2% and 5%, but to reduce the actual spending amount each year by the proportion of their friends and acquaintances who have passed away. This would roughly approximate the optimal decline based on anticipated survival rates.
other words, retirees (on Planet Vulcan) should consume less at older ages. The consumption rate from discretionary wealth is proportional to the survival probability \((t, p_t)\) and is a function of risk aversion, even when the subjective rate of time preferences \((\rho)\) is equal to the interest rate. The rational consumer – planning at age 65 -- is willing to sacrifice some income at the age of 100 in exchange for more income at 80. Stated differently, giving the age of 100 the same preference weight as the age of 80 can only be explained within a lifecycle model if the subjective discount rate \((\rho_t)\) is a time-dependent function that exactly offsets the declining survival probability. It is highly unrealistic for people to have such preferences.

3. The interaction between (longevity) risk aversion and survival probability is quite important. In particular, the impact of risk aversion is to increase the effective probability of survival. So, imagine two retirees with the same amount of initial retirement wealth and pension income (and the same subjective discount rate) but with differing levels of risk aversion \((\gamma)\). The individual with greater risk aversion behaves as if their modal value of life is higher. Specifically they behave as if it is increased by an amount proportional to \(\ln[1/\gamma]\) and they spend less in anticipation of their longer life. Observers will never know if they are longevity-risk averse or just healthier.

4. The optimal trajectory of financial capital (also) declines with age. Moreover, for individuals with pre-existing pension income it is rational to spend-down wealth by some advanced age and live exclusively on the pension income. The wealth depletion time (WDT) can be at age 90 – or even age 80 when the pension income is sufficiently large. Greater (longevity) risk aversion, which is associated with lower consumption, induces greater financial capital at all ages. There is nothing wrong or irrational about planning to deplete wealth by some advanced age.\(^8\)

5. The rational reaction to portfolio shocks (i.e. losses) is non-linear and dependent on when the shock is experienced as well as the amount of pre-existing pension income. One does not reduce portfolio withdrawals by the exact amount of a financial shock unless their risk aversion is \((\gamma = 1)\), a.k.a. Bernoulli utility. For example, if the portfolio suffers an unexpected loss of 30%, the retiree might reduce consumption by only 30% less as a result.

6. Converting some of the initial nest egg into a stream of lifetime income increases consumption at all ages regardless of the cost of the pension annuity. Even when interest rates are low and the cost of $1 of lifetime income is (relatively) high the net effect is that “pensionization” increases consumption. Note that we are careful to distinguish between real world pension

\(^8\) One could therefore say that there are bag-ladies on planet Vulcan.
annuities -- in which the buyer hands over an irreversible sum in exchange for a constant real stream – and Tontine annuities which are at the foundation of most economic models but are completely unavailable in the marketplace.

7. One final result that follows from our analysis, although not pursued in the numerical examples, is counter-intuitive and perhaps even controversial. Borrowing against pension income might be optimal at advanced ages. For individuals with relatively large pre-existing (Defined Benefit) pension income, it might make sense to pre-consume (and enjoy) the pension while they are still young. The lower the (longevity) risk aversion, the more optimal this path becomes.

The “cost” of having a simple analytic expression – described by equations (#1) to (#8) in the appendix -- is that we had to assume a deterministic investment return. Although we used a safe and conservative assumed return for most of the displayed numerical examples, we have for all intents and purposes ignored the last 50 years of portfolio modeling theory. Recall however that our attempt was to shed light on the often-quoted rules of thumb and how they relate to longevity risk, as opposed to developing a full-scale dynamic optimization model.

**Back to Planet Earth**

How might a full stochastic model -- with possible shocks to health and their related expenses as well -- change optimal consumption policies? Assuming agreement on a reasonable model and parameters for long-term portfolio returns, the risk-averse retiree would be exposed to the risk of a negative (early) shock and would plan for this in advance by consuming less. However, with a full menu of investment assets and products available, the retiree would be free to optimize around pension annuities and other downside-protected products, in addition to long-term care (LTC) insurance and other retirement products.

In other words, even the formulation of the problem itself becomes much more complex.

More importantly, the optimal allocation depends on the retiree’s preference for personal consumption versus bequest, as illustrated in Figure #3. A product and asset allocation that is suitable for a consumer with no bequest or legacy motives – those in the lower left-hand corner of the figure – is quite different from the optimal portfolio for someone with strong legacy preferences. This paper assumed the retiree’s objective is to maximize utility of lifetime consumption without any consideration for the value of bequest or legacy.
Also, while some have argued that a behavioral explanation is needed to rationalize the desire for a constant consumption pattern in retirement, we note that very high longevity risk aversion leads to relatively constant spending rates, and might “explain” these fixed rules. In other words, we don’t need a behavioral model to justify constant 4% spending. Extreme risk aversion does it for us.

That said, we are currently working on a sequel in which we derive the optimal portfolio withdrawal rate with pension and tontine annuities in a more robust capital markets environment a la Richard (1975) and Merton (1971). Another fruitful line of research would be to explore the optimal time to retire within the context of a mortality-only lifecycle model, although that also would take us far beyond the current literature.⁹

One thing seems clear: longevity risk-aversion and pension annuities remain very important when giving advice regarding optimal portfolio withdrawal rates. That is the main message of this paper and it does not change on Planet Earth.

Acknowledgements

We thank Zvi Bodie, Larry Kotlikoff, Peng Chen, Francois Gaddene, Mike Zwecher, David Macchia and participants at the Retirement Income Industry Association (RIIA) 2010 conference in Chicago as well as Barry Nalebuff, Sherman Hanna and two anonymous FAJ reviewers for helpful comments. We also acknowledge our colleagues at York University; Pauline Shum, Tom Salisbury, Nabil Tahani, Chris Robinson and David Promislow for helpful discussions on this topic during our research. Finally, we thank Alexandra Macqueen and Faisal Habib at the QWeMA Group (Toronto) for assistance with editing and analytics.

BIBLIOGRAPHY AND REFERENCES


⁹ See Stock and Wise (1990) for an example of this burgeoning literature.


Liu, Hong (2005), Lifetime consumption and investment: Retirement and Constrained Borrowing, working paper, *John M. Olin School of Business*.


**TECHNICAL APPENDIX:**

The value function within the lifecycle model (LCM) during retirement years when labor income is zero, assuming no bequest motive, can be written as follows:

$$\max_c V(c) = \int_0^D e^{-\rho t}(t p_x) u(c_t) dt,$$

(eq.1)

The variable $x$ denotes the current age of the retiree, when the consumption/spending plan is formulated. The parameter $D$, the upper bound of the utility integration, represents the maximum possible lifespan years in retirement. The parameter $\rho$ denotes the subjective discount rate (SDR), a.k.a. personal time preference. The function $(t p_x)$ denotes the conditional probability of survival from retirement age $(x)$ to age $(x + t)$. We parameterize $(t p_x)$ based on the Gompertz law of mortality under which the biological hazard rate is: $\lambda_t = (1/b)e^{(x-m+t)/b}$ which grows exponentially with age. Here $m$ denotes the modal value of life (for example 80 years) and $b$ denotes the dispersion coefficient (for example 10 years) of the future lifetime random variable. Both of these numbers are calibrated to U.S. mortality tables to fit advanced age survival rates.

In our paper the utility function of consumption is assumed to exhibit constant Elasticity of Intertemporal Substitution (EIS), which is synonymous with (and the reciprocal of) constant relative risk aversion (RRA) under conditions of perfect certainty. The exact specification we use is: $u(c) = c^{1-\gamma}/(1 - \gamma)$, where $\gamma$ is the coefficient of relative (longevity) risk aversion which can take on values from Bernoulli ($\gamma=1$) up to possibly infinity.

The actuarial present value function denoted by $a^T_x(v, m, b)$, depends implicitly on the survival probability curve $(t p_x)$ via the parameters $(m, b)$. It is defined and computed using the following:

$$a^T_x(v, m, b) = \int_0^T e^{-vs} (s p_x) ds,$$

(eq.2)

which is the retirement-age “price” – under a real constant discount rate $v$ -- of a life-contingent pension annuity that pays a real $1 per year until the earlier of death and time $T$. Although we don’t include a morality risk premium from the perspective of the insurance company in this valuation model, one could include this by tilting the survival rate towards a longer life.

A closed-form representation of equation (#2) is possible in terms of the incomplete Gamma function $\Gamma(A, B)$, which is available in Excel.
\[
a^T_x(v, m, b) = \frac{br(-v \cdot \exp \left(\frac{x-m}{b} \right)) - br(-v \cdot \exp \left(\frac{x-m+\tau}{b} \right))}{\exp \left((m-x) - v \cdot \exp \left((x-m)/b \right) \right)}.
\]

(eq.2a)

The wealth trajectory (financial capital during retirement) is denoted by \( F_t \) and the dynamic constraint in our model – linked to the objective function equation (#1) -- can now be expressed as follows:

\[
\dot{F}_t = v(t, F_t)F_t - c_t + \pi_0,
\]

(eq.3)

where the dot is shorthand notation for a derivative of wealth (financial capital) with respect to time, \( \pi_0 \) denotes the income (in real dollars) from any pre-existing pension annuities and the function multiplying wealth itself is defined by:

\[
v(t, F_t) = \begin{cases} 
  r, & F_t \geq 0 \\
  R + \lambda_t, & F_t < 0
\end{cases}
\]

(eq.3a)

where \( R \geq r \). The discontinuous function \( v(t, F_t) \) denotes the interest rate on financial capital and allows \( F_t \) to be negative. Credit cards and/or other unsecured lines of credit would be a good example of a situation in which \( v(t, F_t) = R + \lambda_t \). The borrower pays \( R \) plus the insurance (to protect the lender in the event of the borrowers death).

Note that we do not assume a complete liquidity constraint that prohibits borrowing in the sense of Deaton (1991), Leung (1994) or Butler (2001) for example. What we don’t allow is stochastic returns. Equation (#1, #2 and #3) is essentially the Yaari (1965) set-up under which pension annuities are available, but not tontine annuities.

The initial condition is \( F_0 = W \), where \( W \) denotes the investable assets at retirement. The terminal condition is that: \( F_\tau = 0 \), where \( \tau \) denotes the wealth depletion time (WDT) at which point only the pension annuity income is consumed. The existence of a WDT is explored by Leung (1994, 2007) in a series of theoretical papers. In theory the WDT can be at the final horizon time \( \tau = D \), if the pension income is minimal (or zero) and/or the borrowing rate is relatively low. To be very precise here, it is possible for \( F_t < 0 \) for some time \( t < D \). We are not talking about the zero values of the function. Rather, the definition of our WDT is that: \( F_\tau = 0; \forall t > \tau \), which is permanently. One can actually show that when \( R > \rho \) then \( \tau < D \). In our numerical results we assume a high-enough value of \( R \).

The Euler Lagrange Theorem (ELT) from the Calculus of Variations leads to the following. The optimal trajectory \( F_t \) in the region over which it is positive and assuming \( v(t, F_t) = r \), can be expressed as the solution to the following second-order non-homogenous differential equation:
\[ \dot{P}_t - (k_t + r)P_t + rk_tF_t = -\pi_0 k_t, \quad \text{(eq.4)} \]

where the double dots denote the second derivative with respect to time and the time dependent function: \( k_t = (r - \rho - \lambda_t)/\gamma \) is introduced to simplify notation. The real interest rate \( r \) is a positive constant and a pivotal input to the model. Once again we reiterate that equation (4) is only valid until the wealth depletion time \( \tau \). However, one can always force a wealth depletion time \( \tau < D \) by assuming a minimal pension annuity \( \pi_0 = \varepsilon \), as well as a large enough (arbitrary) interest rate \( \nu(t, F_t) \) on borrowing when \( F_t < 0 \). For a more detailed discussion please see Huang and Milevsky (2010).

Moving on, the solution to the differential equation (4) is obtained in two stages. First the optimal consumption rate while \( F_t > 0 \) can be shown to satisfy the equation:

\[ c_t^* = c_0^* e^{k t} (t p_x)^{1/\gamma}, \quad \text{(eq.5)} \]

where \( k = (r - \rho)/\gamma \) and the unknown initial consumption rate \( c_0^* \) will (soon) be solved for. The optimal consumption rate declines when the subjective discount rate \( \rho \) is equal to the interest rate \( r \) and hence \( k = 0 \). This is a very important implication (and observable result) from the lifecycle model. It is rational to plan to reduce one’s standard of living with age, even if \( \rho = r \).

Note also that consumption as defined above includes the pension annuity income \( \pi_0 \). Therefore, the portfolio withdrawal rate (PWR) which is the main item of interest in this paper, is \( (c_t^* - \pi_0)/F_t \) and the initial PWR (a.k.a. retirement spending rate) is \( (c_0^* - \pi_0)/F_0 \).

The optimal financial capital trajectory (also only defined until time \( t < \tau \)) which is the solution to equation (4), can be expressed as a function of \( c_0^* \) as follows:

\[ F_t = \left( W + \frac{\pi_0}{\gamma} \right) e^{r t} - a_x^z(r - k, m^*, b)c_0^* e^{r t} - \frac{\pi_0}{\gamma}, \quad \text{(eq.6)} \]

where the modified modal value in the annuity factor is: \( m^* = m + b \ln \gamma \). The actuarial present value term multiplying time-zero consumption in equation (6) values a life-contingent pension annuity under a shifted modal value of: \( m + b \ln[\gamma] \) and shifted valuation rate of: \( r - (r - \rho)/\gamma \) instead of \( r \). It has no economic interpretation other than being an intermediate step in our solution. However, plugging equation (6) into the differential equation (4) will confirm the solution is correct and valid over the domain: \( t \in (0, \tau) \).
In other words, the value function in equation (#1) -- and hence lifecycle utility -- is maximized when the consumption rate and the wealth trajectory satisfy equation (#5) and (#6) respectively. Of course, these two equations are functions of two unknowns $c_0^*, \tau$, and we now must solve for them. We do this sequentially.

First, from equation (#6) and the definition of the wealth depletion time: $F_\tau = 0$, we can solve for the initial consumption rate:

$$c_0^* = \frac{(W + \pi_0) e^{\pi_0/r}}{a_x^\tau (r - k, m^*, b)} e^{\pi_0/r}. \quad (eq.7)$$

Notice that when $\gamma = 1$ and $\pi = 0$ and $\rho = r$ the entire expression (#7) collapses to $W/a_x^\tau$

Finally, the wealth depletion time $\tau$ is obtained by substituting equation (#7) into equation (#5) and searching the resulting non-linear equation over the range $(0, D)$ for the value of $\tau$ that solves $c_\tau^* - \pi_0 = 0$. In words, if a wealth depletion time exists, then for consumption to remain smooth at that point -- which is actually a foundation of lifecycle theory -- it must converge to $\pi_0$.

Mathematically the wealth depletion time ($\tau$) satisfies the equation:

$$\frac{(W + \pi_0) e^{r\tau - \pi_0}}{a_x^\tau (r - k, m^*, b)} e^{r\tau} (1_P)^{1/\gamma} = \pi_0, \quad (eq.8)$$

In other words:

$$\tau = f(\gamma, \pi_0 | W, \rho, r, x, m, b) \quad (eq.8a)$$

The optimal consumption policy (described by equation #5) and the optimal trajectory of wealth (described by equation #6) are now available explicitly. Practically speaking the wealth depletion time $\tau \leq D$ is extracted from equation (#8) and then the initial consumption rate is obtained from equation (#7). Everything else follows. These expressions can be coded-up in Excel in just a few minutes.
Calibrating Yaari in the 21st Century, Part I: Consumption Under a Stochastic Force of Mortality

H. Huang, M.A. Milevsky\(^1\) and T.S. Salisbury

Draft Version: 1 December 2010

---

\(^1\)Huaxiong Huang is Professor of Mathematics and Statistics at York University. Moshe A. Milevsky is Associate Professor of Finance, York University, and Executive Director of the IFID Centre. Thomas S. Salisbury is Professor of Mathematics and Statistics at York University, all in Toronto, Canada. The contact author (Milevsky) can be reached via email at: milevsky@yorku.ca. The authors acknowledge the helpful comments of seminar participants at the Lifecycle, Insurance, Finance and Economics (LIFE) conference at the Fields Institute. This research is supported in part by grants from NSERC, MITACS and The IFID Centre.
Abstract

We extend the Yaari (1965) framework of consumption over a random-length lifecycle to a stochastic model in which (i.) the force of mortality obeys a diffusion process as opposed to being deterministic, and (ii.) a consumer can adapt their consumption strategy to new information about their mortality rate (a.k.a. health status) as it becomes available. We compare and contrast the optimal consumption paths and carefully examine the impact of mortality rate uncertainty. Our main result is that when utility is logarithmic the initial consumption rate is identical to the Yaari (1965) model even when mortality rates are stochastic. But, in a CRRA framework in which the coefficient of relative risk aversion is greater (smaller) than one the consumption rate is higher (lower) and a stochastic force of mortality changes optimal behavior. Numerical examples are provided in order to gauge the magnitude of this effect. Our results should be relevant to researchers interested in calibrating the lifecycle model as well as those who provide normative guidance (a.k.a. financial advice) to retirees.
1 Introduction and Motivation

Ponder this. Two soon-to-be hypothetical retirees approach a financial economist for guidance on how exactly they should spend their accumulated wealth (a.k.a. nest egg) over their remaining lifetime; a time horizon they both acknowledge is stochastic. Assume both retirees have time-separable and rational preferences and seek to maximize discounted utility of lifetime consumption with the same elasticity of intertemporal substitution ($1/\gamma$), the same subjective discount rate ($\rho$) and the same initial financial capital constraint ($F_0$). They have no declared bequest motives and – for whatever reason – neither are willing (or able) to invest in anything other than a risk-free asset with instantaneous return ($r$); which means they are not looking for guidance on asset allocation. All they want is an optimal consumption plan ($c^*(t); t \geq 0$) guiding them from time zero (retirement) to the last possible time date of death ($t \leq D$). Most importantly, both retirees agree-on and share the same probability-of-survival curve denoted by $p(s)$. In other words they currently live in the same “health state” and the same effective biological age. For example, they both agree on a $p(35) = 5\%$ probability that they survive for 35 years and a $p(20) = 50\%$ probability that they survive for 20 years, etc.

Over 45 years ago, Menahem Yaari (1965) in a classical and highly-cited paper showed exactly how to solve such a problem. He derived the Euler-Lagrange equation for the optimal trajectory of wealth and the related consumption function. He was the first to show how to work with lifetime uncertainty in a lifecycle model (LCM) and amongst other results provided a rigorous foundation for Irving Fisher’s claim that lifetime uncertainty effectively increases consumption impatience and is akin to behavior under higher subjective discount rates.

In Yaari’s (20th century) model both of the above-mentioned retirees would be told to follow identical consumption paths until their random date of death. In fact, they would both be guided to optimally consume $c(t)^* = F(t)/a(t)$, where $a(t)$ is a function of time only and is related to an actuarial annuity factor. We will explain this factor in more detail, later in the paper.

But here is where things get complicated and the impetus for this paper. Although both retirees agree-on the same survival probability curve $p(s)$, they have differing views about the volatility of their health as proxied by a mortality rate volatility. In the language of modern actuarial science, the first retiree (#1) believes that his instantaneous force of mortality (denoted by $\lambda^{DM}(t)$) will grow at a deterministic rate until he eventually dies, while the second retiree (#2) believes that her force of mortality (denoted by $\lambda^{SM}(t)$) will grow at stochastic (but observable) rate until a random date of death. The remaining lifetime random variable is doubly stochastic. While this distinction might sound farfetched and artificial, a number of researchers in the actuarial literature over the last decade or so have advocated for a so-called stochastic force of mortality when pricing and reserving against mortality-
contingent claims. A number of articles – reviewed in Cairns, Blake and Dowd (2006) and a number of papers in the same issue of the *Journal of Risk and Insurance* – have argued that SfM models better reflect the uncertainty inherent in demographic projections *vis a vis* the inability of insurance companies to diversify mortality risk entirely.

When one thinks about it, real-life mortality rates are indeed stochastic, capturing (un-)expected improvements in medical treatment, or (un-)expected epidemics, or even (un-)expected changes to the health status of an individual. Rational consumers choosing to make saving and consumption decisions using models based on deterministic mortality rates would likely agree to re-evaluate those decisions if their views about the values of those mortality rates change dramatically. Our thesis is that decision-making is improved if mortality models reflect the dynamic nature of mortality rates.

We will carefully explain the mathematical distinction between deterministic and stochastic forces of mortality (SfM) in section #2 of this paper, but just to make clear here, at time zero both our hypothetical retirees agree on the survival probability curve $p(s)$. However, at any future time their survival probability curves will deviate from each other depending on the realization of the mortality rate between now and then.

So, motivated by these (21st century) models of mortality, in this paper we derive the optimal consumption function for both retirees; one who believes in – and operates under a – stochastic mortality and one who doesn’t. Stated differently, we will solve the Yaari (1965) model where the optimal consumption plan is given as a function of wealth, time and the evolving mortality rate as a state variable. Indeed, with over 1500 references to the Yaari (1965) paper in the economic literature, and the growing interest in stochastic mortality models in the actuarial community, we believe these results will be of widespread interest.

Recall that in the Yaari model conditioning on the mortality rate was redundant or unnecessary since its evolution over time was deterministic. All one needed was the value of wealth $F(t)$ and time $t$. But, in a stochastic mortality model, the mortality rate itself becomes a state variable. In this paper we show how the uncertainty of mortality interacts with longevity risk aversion ($\gamma$) – which is the reciprocal of the intertemporal elasticity of substitution – to yield an optimal consumption plan.

To briefly preview our results, we describe the conditions under which retiree #1 (deterministic mortality) will start-off consuming more than retiree #2 (stochastic mortality), as well the conditions under which retiree #1 consumes less than retiree #2, and the (surprising) conditions under which they both consume exactly the same. We provide numerical examples under a variety of specific mortality models and examine the magnitude of this effect.

The remainder of this paper is organized as follows. In section #2 we explain in more detail exactly how a stochastic model of mortality differs from the more traditional (and widely used in economics) deterministic force of mortality. In section #3 we take the oppor-
tunity to review Yaari (1965) and set our notation and benchmark for the stochastic model. In section #4 we characterize the optimal consumption function in the stochastic mortality model under the most general assumptions, and prove a theorem regarding the relationship between consumption in this model vs. Yaari (1965). In section #5 we make some specific assumptions regarding the stochastic mortality rate and illustrate the magnitude of this effect, and section #6 summarizes our main results and concludes the paper. The appendix contains mathematical details and algorithms that are not central to our main economic contributions.

First, we explain exactly the difference between deterministic and stochastic force of mortality.

2 Understanding The Force of Mortality

In most of the relevant papers in the LCM literature over the last 45 years – starting with Yaari (1965), Richard (1975), Davies (1985), Leung (1990) – the force of mortality from time zero to the last possible date of death is known with certainty. Ergo, the conditional survival probabilities over the entire retirement horizon are known and predictable at time zero. So, if a 65-year-old retiree is told (by his doctor) that he faces a 5% chance of surviving to age 100 and a 37% chance of surviving to age 90, then by definition there is a 13.5% = (0.05/0.37) probability of surviving to age 100, if he is still alive at age 90. In other words, he makes consumption decisions today that trade-off utility in different states of nature, knowing that if-and-when he reaches the age of 90, there will only be a 13.5% chance he will survive to age 100. In the language of actuarial science, the table of individual \( \{q_{x+i}; i = 0, \ldots, N\} \) mortality rates are known in advance. This is the essence of a deterministic force of mortality and textbook life contingencies. If \( q_{65} \) is the retiree’s probability of dying between age 65 and 66, while \( q_{66} \) is the probability of the same retiree dying between age 66 and 67, then the probability of surviving from age 65 to age 67 is \( (1 - q_{65})(1 - q_{66}) \).

In stark contrast, under a stochastic force of mortality the above predictability – or multiplicative relationship – breaks down. While a 65-year-old might currently face a 5% estimated probability of surviving to age 100 and a 37% chance of reaching age 90, there is absolutely no guarantee that the conditional survival probability from any future age, to age 100 (given the observed mortality rates), will satisfy the ratio. At time zero there is an expectation of what the probability will be at age 90. But, the probability itself is random. This way of thinking – which might be new to economists – is the essence of a stochastic force of mortality and is the impetus for our paper.

Here it is formally. Let \( \lambda(t) \) denote the mortality rate of a cohort of a population, which may be stochastic or deterministic. Let \( \mathcal{F}_t = \sigma(\lambda(q) | q \leq t) \) be the filtration determined
by \( \lambda \). Then individuals in the population have lifetimes of length \( \zeta \) satisfying
\[
P(\zeta > s \mid \zeta > t, \mathcal{F}_\infty) = e^{-\int_t^s \lambda(q) \, dq}.
\] (1)

Assume further that \( \lambda(t) \) is a Markov process, and define the survival function \( p(t, s, \lambda) \) by
\[
p(t, s, \lambda) = E\left[ e^{-\int_t^s \lambda(q) \, dq} \mid \lambda(t) = \lambda \right].
\] (2)

This gives the conditional probability of surviving from time \( t \) to time \( s \), given knowledge of the mortality rate at time \( t \). Therefore
\[
P(\zeta > s \mid \zeta > t, \mathcal{F}_t) = E\left[ e^{-\int_t^s \lambda(q) \, dq} \mid \mathcal{F}_t \right] = p(t, s, \lambda(t)).
\] (3)

If \( t = 0 \) then we write \( p(s, \lambda) \) for \( p(0, s, \lambda) \).

Our basic problem in this paper will be to compare optimal consumption under two models that share a common initial value \( \lambda_0 \) of the mortality rate, as well as a common survival function \( p(t, \lambda_0) \). Typically one will be deterministic and one stochastic. When we do actual computations, we will either choose a specific deterministic model and calibrate a stochastic model to it. Or conversely, we will choose a stochastic model and calibrate the deterministic model to it. Both possibilities are discussed below. It should be clear from the context which model we are discussing. But when it is necessary to make this distinction explicitly, we will write \( \lambda^{\text{DfM}}(t) \) and \( \lambda^{\text{StM}}(t) \).

### 2.1 Deterministic force of Mortality (DfM)

Let \( \lambda_0 = \lambda(0) \) be the initial value of the mortality rate. In the deterministic case,
\[
p(t, \lambda_0) = e^{-\int_0^t \lambda(q) \, dq},
\] (4)
and we can recover \( \lambda(t) \) as \(-p_t(t, \lambda_0)/p(t, \lambda_0)\), where the \( t \)-subscript denotes the time derivative. In other words, if we start with a concrete stochastic model, and obtain the survival curve \( p(t, \lambda_0) \) from it, the above formula determines the calibration of the deterministic force of mortality model to it. This approach is computational simpler, but has the disadvantage that neither the stochastic nor deterministic model is in a simple form, familiar to practitioners. In other words, a “simple” model for the stochastic force of mortality rates leads to a “complicated” model for the deterministic force of mortality, and vice versa.

When doing actual calculations we will start by assuming that \( \lambda(t) \) follows a standard Gompertz model. In other words, that
\[
d\lambda(t) = \eta \lambda(t) \, dt
\] (5)
so \( \lambda(t) = \lambda_0 e^{\eta t} \). The usual form for Gompertz is \( \lambda(t) = b^{-1}e^{(x+t-m)/b} \), so here we are using \( \eta = 1/b \) and \( \lambda_0 = b^{-1}e^{(x-m)/b} \). This model is simple, and takes advantage of long experience calibrating the Gompertz model to real populations.
Note that in the deterministic setting,

\[ p(t, s, \lambda(t)) = e^{-\int_t^s \lambda(q) \, dq} = e^{-\int_t^s \lambda(q) \, dq} / e^{-\int_0^s \lambda(q) \, dq} = p(s, \lambda_0)/p(t, \lambda_0). \]  

(6)

This will typically NOT be true in the stochastic setting. As long as we keep in mind that we are calibrating at time 0 (i.e. to \( p(t, \lambda_0) \) only) that should not cause problems.

<table>
<thead>
<tr>
<th>Table #1: Conditional Survival Probability: Deterministic Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>( x = 65 )</td>
</tr>
<tr>
<td>( x = 70 )</td>
</tr>
<tr>
<td>( x = 75 )</td>
</tr>
<tr>
<td>( x = 80 )</td>
</tr>
<tr>
<td>( x = 85 )</td>
</tr>
<tr>
<td>( x = 90 )</td>
</tr>
<tr>
<td>( x = 95 )</td>
</tr>
<tr>
<td>( x = 100 )</td>
</tr>
<tr>
<td>( \lambda_x )</td>
</tr>
</tbody>
</table>

Table #1 displays a typical deterministic mortality survival probability “matrix” of values together with the corresponding mortality rate at each age \( x \), on the bottom row. Note that these numbers were generated using a (deterministic) Gompertz model in which \( m = 89.335 \) and \( b = 9.5 \). Indeed, given the initial probability of survival from age 65 to any age \( y > 65 \) (which is the first column in Table #1) one can solve for the conditional survival probability from age \( y \) to any age \( z > y \), by dividing the two probability values. This is the essence of equation (6). Alas, when mortality rates are stochastic all numbers \( p(t, s, \lambda(t)) \) beyond the first column in Table #1, are unknown at time zero.

2.2 Stochastic Force of Mortality (SfM)

There are many possible stochastic models to choose from. Starting from the models of Lee and Carter (1992) to a review by Cairns, Blake and Dowd (2006) as well as Wills and Sherris (2010), actuaries have employed a variety of specifications for the stochastic \( \lambda(t) \), subsequently used to price mortality and longevity risk. In what follows in the numerical examples, we adopt a lognormal mortality rate, which is often called the Dothan model for interest rates in the derivative pricing literature. Although it might seem natural to have constant drift and diffusion coefficients, in order to calibrate to a given deterministic model, we allow a time-dependent growth coefficient. For most of the numerical examples provided
later-on we take:
\[ d\lambda(t) = \mu(t)\lambda(t)\, dt + \sigma\lambda(t)\, dB(t) \]  
(7)

where \( B(t) \) is a Brownian motion. This is obviously the source of randomness in the stochastic force of mortality.

There are many ways to select (or calibrate) a stochastic force of mortality to a particular survival curve. The details on how to actually compute this are provided in the second part of the appendix.

With the probability background out of the way, we now review the Yaari (1965) model which is based on a deterministic force of mortality.

### 3 Review of the Yaari (1965) Model

The lifecycle model (LCM) with a random date of death and assuming no bequest motive, can be written as follows:

\[ J = \max_c E \left[ \int_0^D e^{-\rho t} u(c(t)) 1_{\{t \leq \zeta\}} dt \right], \]

where \( \zeta \leq D \) is the remaining lifetime satisfying \( \Pr[\zeta > t] = p(t, \lambda_0) \), defined above in Section #2. When the mortality rate is deterministic one can obviously assume independence between the optimal consumption \( c^*(t) \) and the lifetime indicator variable \( 1_{\{t \leq \zeta\}} \), so that by Fubini’s theorem we can re-write the value function as:

\[ J = \max_c \int_0^D e^{-\rho t} u(c(t)) E[1_{\{t \leq \zeta\}}] dt \]

\[ = \max_c \int_0^D e^{-\rho t} u(c(t)) p(t, \lambda_0) dt. \]

> From this perspective, there really isn’t any more randomness in the model. This is a problem within the calculus of variations subject to some constraints on the function \( c(t) \). The wealth (budget) constraint can be written as:

\[ F(t) = v(t, F(t))F(t) + \pi_0 - c(t), \]

(10)

with boundary conditions \( F(0) = W > 0 \) and \( F(D) = 0 \). The parameter \( \pi_0 \) denotes a constant income rate which we include in this section for comparison with Yaari’s model, but which in subsequent sections will be taken to equal zero; \( c(t) \) is the consumption rate and the control variable in our problem; \( v = v(t, F) \) is the interest rate function defined by:

\[ v(t, F) = \begin{cases} 
  r + \xi \lambda(t), & F \geq 0, \\
  R + \lambda(t), & F < 0,
\end{cases} \]

(11)
where \( r \leq R \leq \infty \). The (new) parameter \( \xi \) indicates and denotes the availability of “actuarial notes” as introduced and described by Yaari (1965). In words, \( v := v(t, F) \) is the investment return \( r \) when wealth is positive so that \( F \geq 0 \). It is equal to the borrowing rate \( R + \lambda(t) \) when wealth is negative, i.e. \( F < 0 \). Note the composite structure which includes the mortality rate \( \lambda(t) \) for an individual at time \( t \) to reflect the reality that unsecured loans are available as long as they are purchased with life insurance protection for the creditor.

When \( \xi = 1 \) and \( R = r \), the function \( v(t, F) \) collapses to \( r + \lambda(t) \) and we are in Case B of Yaari (1965). When \( \xi < 1 \) the investor earns less than their mortality rate when holding actuarial notes, due to adverse selection considerations. Finally, when \( R = \infty \) we are effectively imposing a (no) borrowing constraint and when \( \xi = 0 \) as well we do not allow for holding actuarial notes (or purchasing additional annuities.) For the purposes of this paper and from this point onwards we assume that \( R = \infty \) and \( \xi = 0 \). In a follow-up paper we plan to examine the impact and availability of actuarial notes, i.e. the case when \( \xi > 0 \).

Although this wasn’t explicitly imposed in the Yaari (1965) model, in this paper we operate under a constant relative risk aversion (CRRA) formulation for the utility function. In principle this should mean using \( \bar{u}(c) \), where:

\[
\bar{u}(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma}
\]  

(12)

for \( \gamma > 0 \) and \( \gamma \neq 1 \), with the understanding that when \( \gamma = 1 \) we define \( \bar{u}(c) = \ln[c] \). This family of utilities varies continuously with \( \gamma \).

Of course, it makes no difference to our optimization problems if we shift \( \bar{u} \) by an arbitrary additive constant. So to make scaling relationships easier to express, actual calculations will be carried out using the equivalent utilities

\[
u(c) = \frac{c^{1-\gamma}}{1 - \gamma}
\]  

(13)

for \( \gamma > 0 \) and \( \gamma \neq 1 \). When \( \gamma = 1 \) we take \( u(c) = \bar{u}(c) = \ln[c] \).

The marginal utility of consumption is the derivative of utility with respect to \( c \), which is simply

\[
u_c = c^{-\gamma} > 0.
\]  

(14)

To avoid the distractions of inflation models and assumptions, throughout this paper we assume that the interest rate \( r \) is expressed in real (after-inflation) terms and therefore consumption \( c(t) \) is expressed in real terms as well.

Finally, when \( v(t) = r \) during the entire interval \((0, D)\) then as a consequence of the Euler-Lagrange Theorem, the optimal financial capital trajectory \( F(t) \) must satisfy the following linear second-order non-homogenous differential equation over the values for which \( F(t) \neq 0 \).

\[
F_{tt}(t) - \left(\frac{r - \rho - \lambda(t)}{\gamma}\right) F_t(t) + v \left(\frac{r - \rho - \lambda(t)}{\gamma}\right) F(t) = - \left(\frac{r - \rho - \lambda(t)}{\gamma}\right) \pi_0.
\]  

(15)
When the pension income rate $\pi_0 = 0$ the differential equation collapses to the homogenous case.

### 3.1 Explicit Solution: Gompertz Mortality

When the (deterministic) mortality rate function obeys the (pure) Gompertz law of mortality

$$\lambda(t) = \frac{1}{b} \exp \left( \frac{x + t - m}{b} \right), \quad (16)$$

the survival probability is

$$p(t, \lambda_0) = \exp \left\{ - \int_0^t \lambda(q) \, dq \right\} = \exp \left\{ b\lambda_0(1 - e^{t/b}) \right\}. \quad (17)$$

Here $x$ denotes the age at time 0, $m$ is called the modal value and $b$ is the dispersion coefficient for the Gompertz model. To simplify notation let

$$k(t) = \frac{r - \rho - \lambda(t)}{\gamma}, \quad (18)$$

and recall from the budget constraint that:

$$c(t) = vF(t) - F_t(t) + \pi_0, \quad (19)$$

$$c_t(t) = vF_t(t) - F_{tt}(t). \quad (20)$$

Equation (15) can be rearranged as

$$F_{tt}(t) - vF_t(t) + k(t)(vF(t) - F_t(t)) = -k(t)\pi_0, \quad (21)$$

which then leads to

$$k(t)c^*(t) - c^*_t(t) = 0 \quad (22)$$

The solution to this basic equation is

$$c^*(t) = c^*(0)e^{\int_0^t k(s) \, ds} = c^*(0)e^{\int_0^t \left( \frac{r - \rho - \lambda(s)}{\gamma} \right) \, ds} = c^*(0)e^{(\frac{r - \rho}{\gamma})t - \frac{1}{\gamma} \int_0^t \lambda(s) \, ds} \quad (23)$$

where $c^*(0)$ is the optimal initial consumption rate, to be determined, which is the one free constant resulting from equation (22). Note that when the interest rate $r$ is equal to the subjective discount rate $\rho$, and $\gamma = 1$ (i.e. log utility), the optimal consumption rate at any age $x + t$ is the probability of survival to that age times the initial consumption $c^*(0)$. However, when $\gamma > 1$, which implies higher levels of risk aversion, the optimal consumption rate will decline at a slower rate as the retiree ages. Longevity risk aversion induces people to behave as if they were going to live longer than determined by the actuarial mortality.
rates. We will explore the impact of $\gamma$ on the optimal consumption path in a stochastic force of mortality model, later in Section #4, which is why it’s important to focus on this here.

Mathematically one can see that $(p(t, \lambda_0))^{1/(\gamma+\delta)}$ is greater than $(p(t, \lambda_0))^{1/\gamma}$ for any $\varepsilon > 0$ since $p(t, \lambda_0) < 1$ for all $t$. Finally, note that in the Gompertz mortality model evaluating $(p(t, \lambda_0))^{1/\gamma}$ for a given $(x, m, b)$ triplet is equivalent to evaluating $p(t, \lambda_0)$ under the same $x, b$ values, but assuming that $m^* = m + b \ln \gamma$. This then implies that one can tilt/define a new deterministic mortality rate $\hat{\lambda}_0 = \gamma \lambda_0$ and derive the optimal consumption as if the individual was risk neutral. This will be used later in the explicit expression for $F(t)$ and $c^*(t)$.

Moving on to a solution for $F(t)$, we now substitute the optimal consumption solution (23) into equation (19) to arrive at yet another first-order ODE, but this time for $F(t)$:

$$F_t(t) - \pi_0 + c^*(0) e^{(\frac{\gamma}{\gamma+\delta})t} (p(t, \lambda_0))^{1/\gamma} = 0.$$  \hspace{1cm} (24)

Writing down the canonical solution to this equation leads to:

$$F(t) = \left( \pi_0 \int_0^t e^{-rs} ds - c^*(0) \int_0^t e^{(\frac{\gamma}{\gamma+\delta})s} (p(s, \lambda_0))^{1/\gamma} e^{-rs} ds + F(0) \right) e^{rt},$$  \hspace{1cm} (25)

where $F_0$ denotes the free initial condition from the ODE for $F(t)$ in equation (24). Recall that we still haven’t specified $c^*(0)$, the initial consumption. We will do so (eventually) by using the terminal condition $F(D) = 0$.

To represent the wealth trajectory explicitly define the following (new) Gompertz Present Value (GPV) function

$$a_T^x(r, m, b) = \int_0^T p(s, \lambda_0) e^{-rs} ds = \int_0^T e^{-\int_0^s (r+\lambda(t)) dt} ds$$

$$= \int_0^T e^{-\int_0^s \left( r + \frac{1}{b} \lambda \right) dt} ds$$

$$= b \Gamma(-rb, \exp\left\{ \frac{x-m}{b} \right\}) - b \Gamma(-rb, \exp\left\{ \frac{x-m+T}{b} \right\}) \exp\left\{ (m-x)r - \exp\left\{ \frac{x-m}{b} \right\} \right\}.$$  \hspace{1cm} (26)

The function $a_T^x(r, m, b) = a(t)$ is the age–$x$ cost of a life-contingent annuity that pays $1 per year continuously provided the annuitant is still alive, but only until time $t = T$, which corresponds to age $x + T$. If the individual survives beyond age $(x + T)$ the payout stops. Naturally, when $T = \infty$ the expression collapses to a conventional single premium income annuity (SPIA).

Note that $\Gamma(A, B)$ is the incomplete Gamma function. In other words, equation (26) is analytic and in closed-form.

The reason for introducing the GPV is that combining equation (25) with equation (26) leads to the (very tame looking) expression

$$F(t) = \left( F(0) + \frac{\pi}{r} \right) e^{rt} - a_T^x(r-k, m^*, b) c^*(0) e^{rt} - \frac{\pi_0}{r},$$  \hspace{1cm} (27)
where recall that \( m^* = m + b \ln[\gamma] \). Then, using the boundary condition \( F_\tau = 0 \), where \( \tau \) is the wealth depletion time, we obtain an explicit expression for the initial consumption

\[
c^*(0) = \frac{(F(0) + \pi_0/r) e^{rt} - \pi_0/r}{a^*_T(r - k, m^*, b)e^{rt}}. \tag{28}
\]

### 3.2 Consumption Under DfM: Numerical Examples

In our numerical examples we assume an 86.6\% probability that a 65-year-old will survive to the age of 75, a 57.3\% probability of reaching 85, a 36.9\% probability of reaching 90, a 17.6\% probability of reaching age 95 and a 5\% probability of reaching 100. These are the values generated by the Gompertz law with \( m = 89.335 \) and \( b = 9.5 \). To complete the parameter specifications required for our model, we assume the subjective discount rate \( (\rho) \) is equal to the risk-free rate \( r = 2.5\% \). Within the context of a lifecycle model, this implies that the optimal consumption rates would be constant over time in the absence of longevity and mortality uncertainty.

We are now ready for some results. Assume a 65-year-old with a (standardized) $100 nest egg. Initially we allow for no pension annuity income \( \pi_0 = 0 \) and therefore all consumption must be sourced to the investment portfolio which is earning a deterministic interest rate \( r = 2.5\% \). The financial capital \( F(t) \) must be depleted at the very end of the lifecycle, which is time \( D = (120 - 65) = 55 \) and there are no bequest motives. So, according to equation (28), the optimal consumption rate at retirement age 65 is $4.605 when the risk aversion parameter is \( \gamma = 4 \) and the optimal consumption rate is (higher) $4.121 when the risk aversion parameter is set to (higher) \( \gamma = 8 \).

As the retiree ages \( (t > 0) \) he/she rationally consume less each year – in proportion to the survival probability adjusted for \( \gamma \). For example, in our baseline \( \gamma = 4 \) level of risk aversion, the optimal consumption rate drops from $4.605 at age 65, to $4.544 at age 70 (which is \( t = 5 \)), then $4.442 at age 75 (which is \( t = 10 \)), then $3.591 at age 90 (which is \( t = 25 \)) and $2.177 at age 100 (which is \( t = 35 \)), assuming the retiree is still alive. A lower real interest rate \( (r) \) leads to a reduced optimal consumption/spending rate. All of this can be sourced to equation (23).

Thus, one of the important insights from the Yaari (1965) model is that a fully rational consumer will actually spend less as they progress through retirement. The optimizer spends more at earlier ages and reduces spending with age, even if his/her subjective discount rate (SDR) is equal to (or less than) the real interest rate in the economy.

Intuitively the individual deals with longevity risk by planning to reduce consumption – if that risk materializes – in proportion to the survival probability, linked to their risk aversion. The Yaari (1965) model provides a rigorous foundation to Irving Fisher (1930) statement in his book *Theory of Interest* (page 85): “...The shortness of life thus tends powerfully to increase the degree of impatience or rate of time preference beyond what it
would otherwise be...” and (page 90) “Everyone at some time in his life doubtless changes his degree of impatience for income... When he gets a little older,... he expects to die and he thinks: instead of piling up for the remote future, why shouldn’t I enjoy myself during the few years that remain?”

For additional (case specific) examples of the Yaari (1965) model in the retirement phase, we refer the interested reader to Milevsky and Huang (2010) or a recent paper by Lachance (2010).

3.3 Time-zero Consumption Ratio = Initial Withdrawal Rate

Finally, in the very specific case when \( \pi_0 = 0 \) (which implies that the wealth depletion time is \( \tau = D \)) and the subjective discount rate \( \rho = r \), the retiree must rely exhaustively on his/her initial wealth \( F_0 \). We get

\[
\frac{c^*(0)}{F(0)} = \frac{1}{\alpha_x^D(r - \lambda_0/\gamma, \pi_0, b)}
\]

(29)

We now have all the ingredients to compare with a stochastic model. This ratio is often called the Initial Withdrawal Rate (IWR) amongst financial practitioners and in the retirement spending literature.

4 Optimal Consumption: General Results

In this section we obtain the most general optimal consumption strategy for a retiree maximizing expected discounted utility of consumption, which will include the Yaari (1965) model as a special case. Since our main focus now is on the mortality model, at this stage we make the additional assumption \( \rho = r \), that is, that the subjective discount rate equals the interest rate in the economy. Also, in contrast to the discussion in the previous section, we assume no exogenous pension income, so that \( \pi_0 = 0 \), which then precludes any borrowing. Once again we assume a fixed terminal horizon \( D \), which denotes the last possible date of death. The mathematical formulation is to find

\[
J = \max_{c(s) \text{ adapted}} E \left[ \int_0^T e^{-\int_0^s (r + \lambda(q)) dq} u(c(s)) ds \right| \lambda(0) = \lambda, F(0) = F ].
\]

(30)

Whereas in section #3 of this paper we used basic calculus of variations to derive the optimal trajectory of wealth and the consumption function in the Yaari (1965) model, given the inclusion of mortality as a state variable we must resort to dynamic programming techniques to obtain the optimality conditions. Regardless of the different techniques, we will show how the optimal consumption function collapses to the Yaari (1965) model when the volatility of mortality is zero.
Define:

\[ J(t, \lambda, F) = \max_{c(t)} \text{ adapted } E \left[ \int_t^T e^{-\int_s^t (r+\lambda(q)) dq} u(c(s)) ds \right| \lambda(t) = \lambda, F(t) = F ] \] (31)

As in the deterministic mortality model, the wealth process (which we shall soon see is stochastic) satisfies \( dF(t) = (rF(t) - c(t)) dt \). Assume that there is an optimal control. Then for that control,

\[ E \left[ \int_0^T e^{-\int_t^t (r+\lambda(q)) dq} u(c(s)) ds \right| F_t \] \[ = e^{-\int_t^t (r+\lambda(q)) dq} J(t, \lambda(t), F(t)) + \int_0^t e^{-\int_t^s (r+\lambda(q)) dq} u(c(s)) ds \] (32)

is a martingale. This will likewise give a supermartingale under a general choice of \( c \). Applying Itô’s lemma, we obtain the following Hamilton-Jacobi-Bellman (HJB) equation:

\[ \sup_c \{ u(c) - c J_F \} + J_t - (r + \lambda) J + r F J_F + \mu(t) \lambda J_\lambda + \frac{\sigma^2 \lambda^2}{2} J_{\lambda \lambda} = 0. \] (33)

If there is any possibility of confusion, we will denote this value function \( J_{\text{SM}}(t, \lambda, F) \).

For deterministic mortality, HJB can be obtained by sending \( \sigma \to 0 \) with \( \mu(t) = \eta \), which was equal to \( 1/b \) in the Yaari (1965) model derived in Section #3, as

\[ \sup_c \{ u(c) - c J_F \} + J_t - (r + \lambda) J + r F J_F + \eta \lambda J_\lambda = 0. \] (34)

Moving on to the optimal consumption plan, we solve the HJB equation under CRRA utility as follows: let

\[ u(c) = \frac{c^{1-\gamma}}{1-\gamma}, \quad J = \frac{F^{1-\gamma}}{1-\gamma} a(t, \lambda) \] (35)

and apply the 1st order condition \( c^* = J_F^{-\frac{1}{\gamma}} \). We obtain \( c^* = Fa^{-\frac{1}{\gamma}} \) and get the following equation for \( a(t, \lambda) \):

\[ a_t - (r \gamma + \lambda) a + \gamma a^{1-\frac{1}{\gamma}} + \mu(t) \lambda a_\lambda + \frac{\sigma^2 \lambda^2}{2} a_{\lambda \lambda} = 0 \] (36)

with boundary condition \( a(T, \lambda) = 0 \).

We now solve the PDE for \( a(t, \lambda) \), which we re-write as:

\[ \beta_t + 1 - v \beta + \mu(t) \lambda \beta_\lambda + \frac{\gamma - 1}{2 \beta} \sigma^2 \lambda^2 \beta_\lambda^2 + \frac{1}{2} \sigma^2 \lambda^2 \beta_{\lambda \lambda} = 0. \] (37)

for \( \beta = \beta(t, \lambda) = a(t, \lambda)^{1/\gamma} \). Here also the new variable \( v = r + \frac{\lambda}{\gamma} \). The boundary conditions are \( \beta(T, \lambda) = 0, \beta_\lambda(t, \infty) = 0 \) and at \( \lambda = 0 \) we solve \( \beta_t + 1 - v \beta = 0 \). Note that the optimal consumption rate is \( c = F/\beta \), using shorthand notation.

We are now ready for our main theorem which implicitly answers the question posed in the introduction and motivation to this paper.
4.1 Stochastic Force of Mortality: Main Theorem

Denote by $c^{\text{SfM}}(t, \lambda, F)$ the optimal consumption at time $t$, given $\lambda(t) = \lambda$ and $F(t) = F$, under a stochastic force of mortality (SfM) model. Denote by $c^{\text{DfM}}(t, F)$ the optimal consumption at time $t$, when $F(t) = F$, under a deterministic force of mortality (DfM) model.

**THEOREM**: Assume that the survival functions for the two models agree: $p^{\text{SfM}}(t, \lambda_0) = p^{\text{DfM}}(t, \lambda_0)$ for every $t \geq 0$, and that utility is CRRA$(\gamma)$.

(a) $\gamma > 1 \implies c^{\text{SfM}}(0, \lambda_0, F) \geq c^{\text{DfM}}(0, F)$.

(b) $\gamma = 1 \implies c^{\text{SfM}}(0, \lambda_0, F) = c^{\text{DfM}}(0, F)$.

(c) $0 < \gamma < 1 \implies c^{\text{SfM}}(0, \lambda_0, F) \leq c^{\text{DfM}}(0, F)$.

**PROOF**: To see this, we change point of view, and work exclusively with the stochastic model. So we drop the SfM superscript, and write $p = p^{\text{SfM}}$, $J = J^{\text{SfM}}$, $c^* = c^{\text{SfM}}$, $\lambda = \lambda^{\text{SfM}}$, etc. Within that model, we pose two different optimization problems, depending on the level of information available about $\lambda(t)$. The value function $J(t, \lambda, F)$ solves the problem given before in (31), where $c(t)$ can be any suitable process adapted to $\mathcal{F}_t$. But we define a new value function $J^1(t, F)$ in which we impose an additional constraint on $c(t)$, namely that it be deterministic. More precisely,

$$J(0, \lambda_0, F_0) = \max_{c(s) \text{ adapted}} E \left[ \int_0^T e^{-\int_0^t (r+\lambda(q)) \, dq} u(c(s)) \, ds \right] \quad (38)$$

$$J^1(0, F_0) = \max_{c(s) \text{ deterministic}} E \left[ \int_0^T e^{-\int_0^t (r+\lambda(q)) \, dq} u(c(s)) \, ds \right]$$

$$= \max_{c(s) \text{ deterministic}} \int_0^T e^{-rs} p(s, \lambda_0) u(c(s)) \, ds.$$

We let $c^*$ denote the optimal control for $J$, and $c^1$ denote the optimal control for $J^1$.

Since every deterministic control $c(t)$ is also adapted, we have the basic relationship

$$J(0, \lambda_0, F_0) \geq J^1(0, F_0). \quad (39)$$

On the other hand, the above expression is exactly what the old deterministic model would have given. That is,

$$J^1(0, F_0) = J^{\text{DfM}}(0, F_0) \quad (40)$$

and $c^1 = c^{\text{DfM}}$.

We know from our earlier scaling arguments that $J(t, \lambda, F) = a(t, \lambda) F^{1-\gamma}/(1 - \gamma)$ and $c^*(t, \lambda, F) = a(t, \lambda)^{-1/\gamma} F$ for some function $a \geq 0$. Likewise $J^{\text{DfM}}(t, F) = a_1(t) F^{1-\gamma}/(1 - \gamma)$ and $c^1 = a_1^{-1/\gamma} F$ for some $a_1 \geq 0$. If $\gamma > 1$ then $1 - \gamma < 0$, so $a(0, \lambda_0) \leq a_1(0)$, so $c^* \geq c^1$ at $t = 0$. This shows (a). Likewise if $0 < \gamma < 1$ then $a(0, \lambda_0) \geq a_1(0)$, so $c^* \leq c^1$ at $t = 0$. This shows (c).
Recall that when $\gamma = 1$, we have $u(c) = \ln[c]$. Earlier, when $\gamma \neq 1$, we had $u(c) = c^{1-\gamma}/(1 - \gamma)$ and could make use of a scaling relation. In other words, if $c$ is optimal for $F$, then $kc$ is optimal for $kF$, and that leads to the expression $J(t, \lambda, kF) = k^{1-\gamma}J(t, \lambda, F)$. Or in other words,

$$J(t, \lambda, F) = F^{1-\gamma}J(t, \lambda, 1).$$

With logarithmic utility, the corresponding expression is that

$$J(t, \lambda, kF) = J(t, \lambda, F) + (\ln k) \int_t^T e^{-r(s-t)}p(t, s, \lambda) \, ds.$$  

Or in other words,

$$J(t, \lambda, F) = J(t, \lambda, 1) + (\ln F) \int_t^T e^{-r(s-t)}p(t, s, \lambda) \, ds.$$  

Likewise,

$$J^\text{DIM}(t, F) = J^\text{DIM}(t, 1) + (\ln F) \int_t^T e^{-r(s-t)}p(s, \lambda_0) \, ds.$$  

The first order conditions in the optimization problem then imply that

$$c^* = F/ \int_t^T e^{-r(s-t)}p(t, s, \lambda) \, ds, \quad c^\text{DIM} = F/ \int_t^T e^{-r(s-t)}p(s, \lambda_0)/p(t, \lambda_0) \, ds.$$  

These agree when we send $t \to 0$, showing (b). QED

The theorem certainly proves that $\gamma = 1$ is a point of indifference. The invariance of mortality volatility when utility is logarithmic is reminiscent of similar results in consumption theory where income negates substitution effects. More on this later.

Note that we only use $J^1(0, F)$ above, not $J^1(t, F)$. If we had, we would have had to be careful. The correct definition is that

$$J^1(t, F) = J^\text{DIM}(t, F) = \max_{c(s)} \int_t^T e^{-r(s-t)} \frac{p(s, \lambda_0)}{p(t, \lambda_0)} u(c(s)) \, ds$$

rather than

$$\max_{c(s)} E \left[ \int_t^T e^{-\int_t^s (r + \lambda(q)) \, dq} u(c(s)) \, ds \right] = \max_{c(s)} \int_t^T e^{-r(s-t)} E[p(t, s, \lambda(t))] u(c(s)) \, ds.$$  

These quantities have connections to annuities, as suggested by the fact that the optimal consumption rates given above are, as a fraction of wealth, inverse annuity prices. In particular, $\int_t^T e^{-rs}p(s, \lambda_0) \, ds$ is the (actuarial) price of a deferred annuity, purchased at time 0 with payments starting at time $t$. While $\int_t^T e^{-rs}E[p(t, s, \lambda(t))] \, ds$ is a “forward” annuity price. That is, if at time 0 an insurance company guarantees (a retiree) the right to buy an annuity at time $t$ at a price determined at time 0, then this is that price (computed actuarially, i.e. by discounting mean cash flows).
5 Optimal Consumption: Numerical Examples

We started with a particular survival probability at time zero, namely the Gompertz mortality curve with parameters $m = 89.335$ and $b = 9.5$. The age $x = 65$ survival probabilities to any age $y > x$ are given in Table #1. Both hypothetical retirees agree on these numbers, which means that their initial mortality rate is

$$\lambda_0 = (1/9.5) \exp\{(65 - 89.335)/9.5\} = 0.008125.$$

Over time retiree #1 believes his mortality rate will grow at a rate $\eta = (1/9.5) = 0.10526316$ per year, while retiree #2 believes it will evolve stochastically with a time-dependent growth rate of $\mu(t)$ and a volatility $\sigma$. The actual curve $\mu(t)$ depends on the selected parameter for volatility, since $\mu(t)$ is constrained to match $p(0,\lambda_0)$. The actual process for extracting $\mu(t)$ for any given value of $\sigma$ is rather complicated (although it is not central to our analysis) and is placed in the appendix of this paper. With these numbers in hand – and specifically the function $\mu(t)$ for the drift of the mortality rate – we can proceed to solve the PDEs given in equation (36) and (37), which then lead to the desired optimal consumption function and the initial portfolio withdrawal rate at age 65.

<table>
<thead>
<tr>
<th>Mortality Volatility</th>
<th>$\gamma = 0.5$</th>
<th>$\gamma = 1.0$</th>
<th>$\gamma = 1.5$</th>
<th>$\gamma = 3$</th>
<th>$\gamma = 5$</th>
<th>$\gamma = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma = 0$</td>
<td>7.59%</td>
<td>6.12%</td>
<td>5.58%</td>
<td>5.02%</td>
<td>4.78%</td>
<td>4.61%</td>
</tr>
<tr>
<td>$\sigma = 15%$</td>
<td>7.52%</td>
<td>6.12%</td>
<td>5.60%</td>
<td>5.04%</td>
<td>4.80%</td>
<td>4.62%</td>
</tr>
<tr>
<td>$\sigma = 25%$</td>
<td>7.44%</td>
<td>6.12%</td>
<td>5.62%</td>
<td>5.06%</td>
<td>4.82%</td>
<td>4.63%</td>
</tr>
</tbody>
</table>

Notes: Retirement age 65, interest rate $r = 2.5\%$, mortality $\lambda_0 = 0.0081$

Table #2 provides a variety of numerical examples across different values of (mortality volatility) $\sigma$ and (risk aversion) $\gamma$, once again assuming that the retirees are both at age $x = 65$ with observable mortality rate $\lambda_0 = 0.0081$. As we proved in section #3, and discussed above, the consumption rate is the same across all levels of mortality volatility when $\gamma = 1$. It increases relative to DfM when $\gamma > 1$ and decreases relative to DfM when $\gamma < 1$. Notice that impact of stochastic mortality on optimal withdrawal rates is reduce as the value of risk aversion increases. Notice how at a coefficient of relative risk aversion $\gamma = 10$, the portfolio withdrawal rates are (approximately) 4.6%.

Note that the $\sigma$ values provided are rather ad hoc and have not been estimated from any particular demographic dataset. We refer the interested reader to the paper by Bauer, et. al. (2008) for an empirical discussion around the estimation of the volatility of mortality. Rather, our objective here is to explore whether or not volatility has a noticeable impact on behavior.
6 Discussion and Conclusion:

In this article we extended the classical Yaari (1965) model of consumption over a random-length lifecycle, to a model in which individuals can adapt behavior to new information about mortality rates which is a proxy for changes in health status. Yaari (1965) was the first to include lifetime uncertainty in a Ramsey-Modigliani lifecycle model and amongst other results, he provided a rigorous foundation for Irving Fisher’s claim that lifetime uncertainty increases consumption impatience and is akin to higher subjective discount rates. Yaari (1965) is at the core of most (classical) macro-economic models used within a lifecycle framework, but to our knowledge his work not been extended into the realm of 21st century models of mortality and longevity risk.

We built this extension by assuming that (i) the instantaneous force of mortality is stochastic and obeys a diffusion process as opposed to being deterministic, and (ii) that a utility-maximizing consumer can adapt their consumption strategy to new information about their mortality rate (a.k.a. current health status) as it becomes available. Our diffusion model for the stochastic force of mortality was quite general and the models were inspired by (a.k.a. borrowed from) the recent literature in the actuarial science arena which has argued for stochasticity in mortality rates.

We focused our attention on the “retirement income” stage of the lifecycle model (LCM) where health considerations are likely to be more prevalent and to avoid complications induced by wages, labor and human capital consideration. Thus, we refer to our general consumer as the retiree and the retirement consumption rate $c(t)$ scaled by wealth $W$, as the initial withdrawal rate.

In most\(^1\) of the papers in the continuous-time LCM literature the force of mortality from time zero to the last possible date of death is known with certainty at time zero. Ergo, the conditional survival probabilities over the entire retirement horizon are known and predictable. So, if a 65-year-old retiree is told (by his doctor) that he faces a 5% chance of surviving to age 100 and a 37% chance of surviving to age 90, then by definition there is a $13.5\% = (0.05/0.37)$ probability of surviving to age 100, if he is still alive at age 90. In other words, he makes consumption decisions today that trade-off utility in different states of nature, knowing that if-and-when he reaches the age of 90, there will only be a 13.5% chance he will survive to age 100. In the language of actuarial science, the table of $q_x$ mortality factors are known in advance. This is the essence of a deterministic force of mortality – which is textbook life contingencies – and is at the core of most economic lifecycle models such as Richard (1975), Davies (1981), Leung (1990) or Lachance (2010) that model consumption

---

\(^1\)An exception would be the paper by Cocco and Gomes (2009) who recognize the stochasticity of mortality and use a Lee-Carter type model to numerically calibrate a lifecycle model. While our paper is motivated by the same consideration, our objective here is to extend the Yaari (1965) model, while retaining the essence of his other non-mortality assumptions.
and portfolio choice problems with lifetime uncertainty.

In stark contrast, under a stochastic force of mortality the above predictability simply breaks down. While a 65-year-old might currently face a 5% chance of surviving to age 100 and a 37% chance of reaching age 90, there is absolutely no guarantee that the conditional survival probability from any future age, to age 100 (given the observed mortality rates), will satisfy the ratio. At time zero there is an expectation of what the probability will be at age 90. But, the probability itself is random. This way of thinking – which might be new to economists – is the essence of a stochastic force of mortality. The actuarial science literature has virtually exploded in this area, but so far it has had little impact on the economics (consumption theory) literature.

So, to phrase our question differently: how does an increase in the information set and the ability to act on this information change the optimal consumption function from the Yaari (1965) model?

To compare results with previous literature and set notation, in the first part of this paper we re-derived the optimal consumption function – under a deterministic force of mortality (DFM) using techniques from the calculus of variations. We provided a closed-form expression for the entire consumption rate function under a Gompertz mortality assumption. With those benchmark results in place, we derived the optimal consumption strategy under a stochastic force of mortality (SfM), by expressing and solving the relevant Hamilton-Jacobi-Bellman (HJB) equation. In addition to the time variable, two state variables in the resulting PDE are current wealth and the current mortality rate. The methodology and mechanics were explained carefully and in great detail in section #4.

This basic set-up enabled us to extract a number of insights regarding the impact of stochastic mortality rates on optimal consumption in general and retirement spending rates in particular. Indeed, framing our model within the context of a retiree is especially relevant in the 21st century. A greater fraction of the population is retiring without access to defined benefit (DB) pension income, is faced with increased mortality (rate volatility) and health uncertainty and must personally decide how to spend their accumulated wealth over their remaining lifecycle.

Although a variety of different authors – most recently Cocco and Gomes (2009) – have explored similar problems, here is a summary of results and insights we believe are new to the literature.

1. As one might expect, the ability to adapt consumption to information about health status and unexpected changes in mortality rates is welfare enhancing. The value function – i.e. discounted utility of lifetime consumption – is uniformly higher at time zero, even though the trajectory of consumption in a SfM model is non-smooth. Recall that a by-product of Yaari (1965) is a smooth consumption path.

2. Retirees with (i) no bequest motives, (ii) constant relative risk aversion (CRRA) pref-
erences, and (iii) subjective discount rates equal to the interest rate are expected to consume less as they age since they prefer to allocate consumption into states of nature where they are most likely to be alive. This is the conventional diminishing marginal utility argument. In our model, while this is not true in every state of nature, it is true on average. In particular, a positive shock to the mortality rate in the form of pleasant health news (perhaps a cure for cancer) will reduce consumption instantaneously and further than expected at time zero, and a negative shock to the mortality rate (for example, being diagnosed with terminal cancer) will increase consumption beyond what was expected.

3. Our particular representation enables researchers to easily compare the consumption strategy of retirees who can react to changes in mortality rates and health status, with retirees who must set their plans in stone at time zero and cannot adapt to (or might even be ignorant of) new information. With proper calibration, this allows us to run a stochastic vs. deterministic horse race.

4. When the coefficient of CRRA (denoted by $\gamma$) is equal to one, and the retiree has logarithmic utility preferences, the optimal consumption rate at time zero is identical in both models. In other words, a retiree who cannot adjust their consumption plan as mortality rates evolve starts-off with the exact same consumption rate as the (more knowledgeable) consumer who can adapt to changes in mortality rates and health status. Although the path of their respective consumption will diverge over time – depending on the evolution of mortality rates – initially they are the same. We found this result to be most interesting.

5. In contrast, when the coefficient of CRRA is greater than one and the retiree is more risk-averse compared to a logarithmic utility maximizer, the initial consumption rate is higher in the stochastic model vs. the deterministic model. In other words, as one might expect the ability to adapt to changes in health status and new information about mortality rates allows the retiree to be more generous at time zero – relative to their “information ignorant” neighbor.

6. When the coefficient of the CRRA is between one and zero, which is at the razor’s edge of longevity risk neutrality, the result is reversed. The canonical retiree in a stochastic mortality model will consume less compared to their neighbor who is operating under deterministic mortality assumptions. Once again, the switch-over point is $\text{CRRA} = 1$, logarithmic utility. Not withstanding the above results, the absolute consumption rate at time zero is uniformly higher the lower the coefficient of relative risk aversion, which is identical to the Yaari (1965). This is a manifestation of longevity risk aversion. The retiree is concerned about living a long time, and therefore consumes less today to protect themselves and self-insure consumption in old age.
7. The calibration of our economic model leads to some interesting by-products in actuarial finance. In particular, in order to construct a stochastic force of mortality that matches or fits a pre-determined Gompertz survival curve – the most popular and frequently used analytic law in this literature – one requires a lognormal diffusion process in which the drift itself grows even faster than exponentially over time. This can be tricky to navigate and caution is required when solving the optimal consumption PDE derived and presented in Section #4. The numerical technique requires us to move backwards thru a space-time grid in which one of the main parameter values in (explosively) large. Likewise, if one assumes a constant coefficient lognormal model, which is a geometric Brownian motion (GBM) of mortality, the resulting survival curve exhibits tails which are quite thicker than the Gompertz model, for the same life expectancy value. Either way, selecting diffusion and drift rate parameters that match a mortality table in expectation lead to large biases in the tail probabilities. We provide an alternative formulation in the appendix.

8. Depending on the actual model selected for the time-zero survival probability, for example Gompertz, or exponentially distributed, or logistic, the changes in the optimal consumption rates can range from as little as 1% to as large as 10% depending on initial age, volatility of mortality, and the actual value of the coefficient of relative risk aversion. In section #5 we offered a number of case-specific insights relating the magnitude of the change in consumption in a stochastic vs. deterministic model. It would therefore be interesting to see if actual consumer behavior as measured in some of the longitudinal databases, can be better described by a stochastic mortality representation.

Moving forward, there are a number of avenues to pursue in future research. A natural extension to this kind of thinking would be to explore the impact of stochastic investment returns as well as mortality rates and thus include a strategic asset allocation dimension. Another item on the research agenda – and one that is occupying us currently – is to dig further into this framework and explore the role of health and mortality-contingent claims in stochastic mortality model. Recall that one of Yaari’s (1965) noted results is that lifecycle consumers with no bequest motives should hold all of their wealth in actuarial notes. These are essentially pension annuities in a deterministic mortality model. However, in the presence of a stochastic mortality, it is no longer clear how an insurance company would price pension annuities, given the systematic risk involved. In such a model, a retiree would have to choose between investing wealth in a tontine pool, with corresponding stochastic returns or purchasing a pension annuity with a deterministic consumption flow, but possibly paying a risk-premium for the privilege. In fact, the optimal portfolio allocation might be a mixture between tontines and annuities. Alas, we leave this work for a subsequent paper.
References


7 Appendix

In this appendix will fill-in some of the missing details on how to calibrate, measure and fit a stochastic mortality model to a given probability curve so that the initial survival rates are identical.

7.1 Matching Time-Zero Survival Curves

Given a deterministic model (Gompertz in our numerical examples), we can compute the time-zero survival function \( p(t; 0) \). We plan to match this using a stochastic model, by a suitable choice of parameters. This means that at time 0 the two models deliver identical survival probabilities. Recall that at times other than \( t = 0 \) the comparison will no longer be meaningful, even controlling for the current observed mortality rate, because the mismatch between conditional survival probabilities means that the two models give different views of lifetimes going forward.

Let \( (t) = e^{-\int_0^t \lambda(s) ds} \) and define a pseudo-density \( q(t; \lambda) \) by the formula

\[
E[ \Lambda(t) \phi(\lambda(t)) ] = \int_0^\infty \phi(\lambda)q(t, \lambda) \, d\lambda.
\]

Then \( p(t, \lambda_0) = \int_0^\infty q(t, \lambda) \, d\lambda \). By Itô’s lemma,

\[
\phi(\lambda(t)) \Lambda(t) = \phi(\lambda_0) + \int_0^t \Lambda(s) \left[ \mu(s)\lambda(s)\phi'(\lambda(s)) + \frac{1}{2} \sigma^2 \lambda(s)^2 \phi''(\lambda(s)) - \lambda(s)\phi(\lambda(s)) \right] ds
\]

\[
+ \int_0^t \Lambda(s)\mu(s)\lambda(s)\phi'(\lambda(s)) \, dB(s).
\]

Take expectations and differentiate with respect to \( t \). We get

\[
\int_0^\infty \phi(\lambda)q_t(t, \lambda) \, d\lambda = \int_0^\infty \left[ \mu(t)\lambda\phi'(\lambda) + \frac{1}{2} \sigma^2 \lambda^2 \phi''(\lambda) - \lambda\phi(\lambda) \right] q(t, \lambda) \, d\lambda
\]

with initial condition \( q(0, \cdot) = \delta_{\lambda_0} \). Using integration by parts (for \( \phi \) vanishing fast at 0 and \( \infty \)), we have

\[
q_t(t, \lambda) = -\mu(t) \frac{\partial}{\partial \lambda} [\lambda q(t, \lambda)] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \lambda^2} [\lambda^2 q(t, \lambda)] - \lambda q(t, \lambda).
\]

So if \( \mu(t) \) is known for \( 0 \leq t \leq t_1 \), then all expectations \( \int_0^\infty q(t_1, \lambda)\phi(\lambda) \, d\lambda \) can be found by solving the forward equation for \( q \) and then integrating against \( \phi \).

Let \( \lambda_{(1)} = \int_0^\infty \lambda q d\lambda \) and \( \lambda_{(2)} = \int_0^\infty \lambda^2 q d\lambda \) be the first two moments of \( q(t, \lambda) \). Note that the zeroth moment is the survival probability, so we can integrate (by parts) the forward PDE for \( q \) and the product of \( \lambda \) and the forward PDE and obtain the following relationships

\[
\lambda_{(1)} = -\frac{dp}{dt},
\]

\[
\lambda_{(2)} = \mu(t)\lambda_{(1)} - \frac{d\lambda_{(1)}}{dt}.
\]
Combined the two expressions, we have
\[ \mu(t) = \frac{d\lambda(1)}{dt} + \lambda(2). \]  
(52)

Replacing \( \mu(t) \) in the forward PDE for \( q \) and obtain an integro-differential equation
\[ q_t(t, \lambda) = \frac{-d\lambda(1)}{\lambda(1)} \frac{\partial}{\partial \lambda} \left[ \lambda q(t, \lambda) \right] + \sigma^2 \frac{\partial^2}{\partial \lambda^2} \left[ \lambda^2 q(t, \lambda) \right] - \lambda q(t, \lambda), \]  
(53)

or
\[ q_t(t, \lambda) = \frac{-\sigma^2}{\eta^2} + \int_0^\infty \lambda^2 q(t, \lambda) d\lambda \frac{\partial}{\partial \lambda} \left[ \lambda q(t, \lambda) \right] + \sigma^2 \frac{\partial^2}{\partial \lambda^2} \left[ \lambda^2 q(t, \lambda) \right] - \lambda q(t, \lambda), \]  
(54)

which we can solve numerically with the initial condition \( q(0, \lambda) = \delta(\lambda - \lambda_0) \).

We solve the integro-differential equation for \( q \) numerically first, obtain the value of \( \mu(t) \). We then solve the HJB equation for optimal consumption as before, with the constant \( \mu \) now replaced by the function \( \mu(t) \) at \( \lambda = 0 \).

Finally, we should record a couple of remarks about the form of \( \mu(t) \). First of all,
\[ \mu(0) = \eta. \]  
(55)

To see this, observe that \( p_t(0, \lambda_0) = -E[\lambda_0 \Delta_0] = -\lambda_0. \) So \( \lambda_0^2 = E[\lambda_0^2 \Delta_0^2] = p_{tt}(0, \lambda_0) + \lambda_0 \mu(0). \) But \( p_{tt}(0, \lambda_0) \) can be computed explicitly, since it is Gompertz, to give \( \lambda_0^2 - \lambda_0 \eta. \) This implies that \( \mu(0) = \eta \).

Second, note that \( \mu(t) \) should be increasing in \( \sigma \). The mean \( E[e^{-\int_t^s \lambda(q) dq}] \) doesn’t change with \( \sigma \), so by convexity of the exponential, the median of this quantity must decrease as we increase the variance. In other words, \( \mu(t) \) must rise. Put another way, this expectation is driven by the possibility of relatively larger values of the exponent, ie of abnormally low values of \( \lambda \). As \( \sigma \) rises, the impact of longevity risk gets more pronounced, and to compensate for that the growth rate \( \mu(t) \) must also rise.

### 7.2 Evolution of the Probability

Recall that we have been denoting conditional survival probabilities as
\[ E[e^{-\int_t^s \lambda(q) dq} | \mathcal{F}_t] = p(t, s, \lambda(t)) \]  
(56)

We wish to compute the following quantity:
\[ \Pr \left( \frac{p(t, s, \lambda(t))}{p(t, \lambda_0)} > \frac{p(s, \lambda_0)}{p(t, \lambda_0)} \right). \]  
(57)

We carry this out in several steps:
• Calibrate the models, to obtain the growth function $\mu(t)$.

• Compute a table of Gompertz survival probabilities $G(t, s) = p(s, \lambda_0)/p(t, \lambda_0)$, for the various values $s > t$ of interest.

• For each fixed $s$ of interest, solve a backward equation to find the function $p(\cdot, s, \cdot)$.

• For each $t$ of interest, compute the inverse function $p_{ts}$ of $p(t, s, \cdot)$. More precisely, compute the value $p_{ts}(G(t, s))$ such that $p(t, s, \lambda) > G(t, s) \Leftrightarrow \lambda < p_{ts}(G(t, s))$.

• Finally, use the lognormal distribution of $\lambda(t)$ to compute the desired probability. We have $\lambda(t) = \lambda_0 \exp(\sigma B(t) + \int_0^t (\mu(q) - \frac{1}{2}\sigma^2) dq)$, so the probability we want is just

$$
\Phi \left( \frac{1}{\sigma \sqrt{t}} \left[ \ln \left( \frac{p_{ts}(G(t, s))}{\lambda_0} \right) - \int_0^t (\mu(q) - \frac{1}{2}\sigma^2) dq \right] \right) \tag{58}
$$

The backward equation mentioned above is completely standard. For $t \leq v \leq s$ we know that

$$
E \left[ e^{-\int_v^s \lambda(q) dq} \mid \mathcal{F}_v \right] = e^{-\int_v^s \lambda(q) dq} p(v, s, \lambda_v) \tag{59}
$$
is a martingale. So $p(v, s, \lambda)$ satisfies

$$
p_v + \mu(v) \lambda p_\lambda + \frac{1}{2} \sigma^2 \lambda^2 p_{\lambda\lambda} - \lambda p = 0 \tag{60}
$$
with terminal condition $p(s, s, \lambda) = 1$ and boundary conditions $p(v, s, 0) = 1$ and $p(v, s, \infty) = 0$.

### 7.3 An Alternate Model for Mortality Rates

In the body of the paper itself (and for the numerical examples) we employed a LogNormal specification for the mortality rate $\lambda(s)$, in which the volatility of mortality was constant $\sigma$, and the drift rate (function) was iterated to fit a given exogenous survival probability curve. This exhibited certain numerical instabilities so as an alternate model we now consider Cox-Ingersoll-Ross (square root diffusion) type dynamics:

$$
d\lambda(t) = [\theta(t) \lambda_0 + \nu \lambda(t)] dt + \sigma \lambda(t)^{1/2} dB(t) \tag{61}
\lambda(0) = \lambda_0.
$$

In analogy with affine term structure models, we seek functions $A(t, s)$ and $B(t, s)$ such that $p(t, s, \lambda) = e^{A(t, s)-\lambda B(t, s)}$. Therefore $e^{-\int_0^t \lambda(q) dq} p(t, s, \lambda_t)$ is a martingale in $t \leq s$, for each fixed $s > 0$. Applying Itô’s lemma, we conclude that

$$
A_t - \lambda B_t - [\theta \lambda_0 + \nu \lambda] B + \frac{\sigma^2 B^2}{2} \lambda - \lambda = 0. \tag{62}
$$
It follows that $A_t = \theta \lambda_0 B$ and $B_t + \nu B - \frac{\sigma^2 B^2}{2} + 1 = 0$, with terminal conditions $A(s, s) = 0$ and $B(s, s) = 0$. Since the ODE for $B$ is autonomous, we conclude that $B(t, s) = \psi(s-t)$ where $\psi' - \nu \psi + \frac{\sigma^2 \psi^2}{2} - 1 = 0$, $\psi(0) = 0$. As in the CIR model, this equation can be solved explicitly, giving

$$\psi(t) = \frac{2(e^{\gamma t} - 1)}{(h - \nu)(e^{\gamma t} - 1) + 2h}$$

(63)

where $h = \sqrt{\nu^2 + 2\sigma^2}$. We also have $A(t, s) = \int_t^s \theta(q) \lambda_0 B(q, s) dq$, and therefore

$$e^{-\frac{\lambda_0}{h}(e^\gamma s - 1)} = p(0, s, \lambda_0) = e^{\lambda_0} \left[ \int_0^s \psi(s-q)\theta(q) dq - \psi(s) \right].$$

(64)

In other words, we calibrate this model by fixing parameters $\nu$ and $\sigma$, and then letting $\theta(s)$ solve the integral equation

$$\int_0^s \psi(s-q)\theta(q) dq = \psi(s) - \frac{1}{\eta} (e^{\gamma s} - 1).$$

(65)

Because $\psi(0) = 0$, the integral equation will be problematic for small $s$. So we fill in the values of $\psi(s)$ for small $s$ by expanding in a power series. The series solution for $\psi$ is

$$\psi(s) = \sum_{n=1}^{\infty} \psi_n s^n, \quad \psi_1 = 1, \quad \psi_2 = \frac{\nu}{2}$$

$$\psi_{n+1} = \frac{1}{n+1} \left[ \nu \psi_n - \frac{\sigma^2}{2} \sum_{j=1}^{n-1} \psi_j \psi_{n-j} \right], \quad n \geq 2.$$  

(66)

If $\theta(q) = \sum_{n=0}^{\infty} \theta_n q^n$ then the LHS of (65) is

$$\sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \theta_j \psi_k \int_0^s (s-q)^k q^j dq = \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \theta_j \psi_k s^{j+k+1} \int_0^1 (1-u)^k u^j du$$

$$= \sum_{n=2}^{\infty} s^n \sum_{k=1}^{n-1} \theta_{n-k-1} \psi_k \chi_{n,k}$$

where (by the binomial theorem)

$$\chi_{n,k} = \sum_{i=0}^{k} \binom{k}{i} (-1)^i \frac{1}{i + n - k}.$$  

(68)

Equating, we obtain that $\theta_0 = \nu - \eta$, and for $n \geq 3$

$$\theta_{n-2} = \frac{1}{\psi_1 \chi_{n,1}} \left[ \psi_n - \frac{\eta^{n-1}}{n!} - \sum_{k=2}^{n-1} \theta_{n-k-1} \psi_k \chi_{n,k} \right].$$

(69)
The first few terms are
\[
\begin{align*}
\psi_1 &= 1, & \psi_2 &= \frac{\nu}{2}, & \psi_3 &= \frac{\nu^2 - \sigma^2}{6}, & \psi_4 &= \frac{\nu^3 - 4\sigma^2\nu}{24}, \\
\chi_{3,1} &= \frac{1}{6}, & \chi_{3,2} &= \frac{1}{3}, & \chi_{4,1} &= \frac{1}{12}, & \chi_{4,2} &= \frac{1}{12}, & \chi_{4,3} &= \frac{1}{4},
\end{align*}
\]
(70)
\[
\theta_0 = \nu - \eta,
\]
\[
\theta_1 = \frac{1}{\chi_{3,1}} \left[ \psi_3 - \frac{\eta^2}{6} - \theta_0 \psi_2 \chi_{3,2} \right] = \eta (\nu - \eta) - \sigma^2,
\]
\[
\theta_2 = \frac{1}{\chi_{4,1}} \left[ \psi_4 - \frac{\eta^3}{24} - \theta_1 \psi_2 \chi_{4,2} - \theta_0 \psi_3 \chi_{4,3} \right] = \eta^2 (\nu - \eta) - \frac{3\sigma^2 \nu}{2}.
\]

7.3.1 Comparing with the LogNormal model

To compare with the LogNormal model, we also investigated the case where \( \theta = 0 \) while letting \( \nu \) be a function of time. In this case, we arrive at the following problem

\[
\frac{\partial B}{\partial t} + \nu(t)B - \frac{\sigma^2}{2}B^2 + 1 = 0,
\]

with

\[
B(s, s) = 0, \quad B(0, s) = \frac{1}{\eta} (e^{\eta s} - 1).
\]

We solve this problem numerically for \( \nu(t) \) using the following algorithm.

1. First we discretize the time domain \([0, s]\) with a grid \( s_j \), with \( s_0 = 0, \ldots, s_n = s \);

2. We discretize the time domain \([0, t]\) with the grid \( t_i \) with the same time step size, for \( t \leq s \);

3. We denote \( B_{i,j} \) as the approximation of \( B(t_i, s_j) \), \( \nu_i \) as the approximation of \( \nu(t_i) \);

4. We discretize the first order differential equation for \( B \) as

\[
\frac{B_{i,j} - B_{i-1,j}}{\delta t} + \nu_{i-1} B_{i-1,j} - \frac{\sigma^2}{2} B_{i-1,j}^2 + 1 = 0
\]

and use it to find \( B_{i,j} \) for \( i = 0, \ldots, j - 1 \), with initial condition \( B_{0,j} = B(0, s_j) \);

5. When \( i = j \), we apply the other boundary condition \( B_{j,j} = B(s_j, s_j) = 0 \) and use

\[
\frac{B_{j,j} - B_{j-1,j}}{\delta t} + \nu_{j-1} B_{j-1,j} - \frac{\sigma^2}{2} B_{j-1,j}^2 + 1 = 0
\]

to find the value of \( \nu_{j-1} \) as

\[
\nu_{j-1} = \frac{\sigma^2}{2} B_{j-1,j} + \frac{1}{\delta t} - \frac{1}{B_{j-1,j}}
\]

for \( j = 2, \ldots, n \). Note that when \( j = 2 \), we do need to solve equation (71).