Literature Review

Longevity Risk and Retirement Income Planning

Patrick J. Collins
Huy D. Lam
Josh Stampfli

CFA Institute
Research Foundation

Draft: May 2015
Accepted For Publication July 2015
INTRODUCTION

Over the last fifty years, there has been an accretion of research on the topic of longevity risk and retirement income planning from scholars and practitioners in diverse fields:

- **Actuaries** are interested in the factors that determine pricing of contracts guaranteeing lifetime income;

- **Financial Economists** are interested in building models reflecting the evolution of retirement portfolios under the stress of expenses and withdrawals, and on using models to optimize outcomes expressed in both dollar-wealth and utility terms;

- **Investment Advisors** are interested in how best to advise clients on a variety of retirement and intergenerational wealth management issues;

- **Trustees** charged with providing lifetime income to current beneficiaries and terminal wealth to remaindermen are interested in how to fulfill their fiduciary duties prudently and impartially; and,

- **Investors** are interested in how much money they can safely spend, gift or bequeath from their retirement portfolio.

Not only is the volume of research vast, but the range of publications reflects pedagogy from academic fields that traditionally have little overlap in readership.

Although hundreds of authors have contributed to our understanding of how retirement income portfolios behave in the presence of distributions, we limit the focus primarily, although not exclusively, to academic studies. This not because the numerous articles appearing in trade-oriented publications lack value; rather because the academic literature is, for the most part, free of sales and marketing agendas. As such, it represents an independent, objective source of credible information. A more comprehensive and technically detailed literature survey, arranged in chronological order, is available in “An Annotated Bibliography: Longevity Risk and Portfolio Sustainability [1965 – 2014].” The document—an approximately 380 page resource—is available at


It comments extensively on both academic and practitioner-oriented literature.
Unlike the aforementioned Annotated Bibliography, this review is organized both chronologically and thematically. Its focus is on research studies that model and evaluate portfolio design, implementation and management strategies for producing adequate lifetime income for retired investors. Unless otherwise indicated, it does not consider wealth accumulation during the investor’s pre-retirement working years. Where appropriate, the survey also extends to literature that considers gifting and bequest objectives.


Any summary of mathematically complex articles must inevitably result in oversimplifications which, to the authors of such articles, may seem to distort their work. Although we offer no defense against this charge other than ‘good faith,’ nevertheless we suspect that the benefit of presenting, in relatively non-technical language, a broad survey of research articles outweighs any embedded faults. Readers wishing to explore the mathematics underlying academic research will find a greater level of technical explication in the aforementioned Annotated Bibliography. Of course, there is no substitute for reading the original research.

Research conclusions are byproducts of mathematical models. Although academic insights derived from these models influence economic theory they are less likely to contribute to normative economics. Research models often incorporate both simplifying assumptions and mathematical approaches in order to achieve tractability. This observation simply acknowledges that models yield benefits—insights into how critical variables interact—and manifest limitations. A research study’s conclusions do not necessary translate into investment planning prescriptions for actual investors.

¹ The Society of Actuaries sponsors the first three studies. They are on the Society’s website (http://www.soa.org).
THEMES AND STRUCTURE OF THE LITERATURE REVIEW

Longevity Risk
Longevity risk is the likelihood that an investor will outlive a portfolio of financial assets tasked with providing lifetime income. The U.S. Social Security Administration’s general population mortality table evidences an increase in life expectancy over the course of the last century. Within the blue-collar population group there is an emerging and well-documented retirement crisis. Primary economic reasons encompass a combination of factors including the decline of defined benefit pension plans, a discomfort with investing in higher expected return assets—a variation on a financial illiteracy explanation—and, stagnation in real wage growth leading to difficulty in accumulating pre-retirement wealth. Within the high-income, white-collar population group it is not uncommon for one or more spouses to live beyond age 90. This means that an investment portfolio may have a planning horizon greater than 35 years assuming an investor in their mid-60s. Additionally, this population group often has a high standard of living which may require substantial cash flows to continue unabated throughout their lifespan. In a low-yield environment, traditional fixed income-oriented strategies for providing adequate and sustainable income are under pressure due to the desire to maintain established lifestyles.

Actuarial Solutions vs. Investment Solutions
Some authors assert that capital market volatility calls into question the viability of using risky-asset portfolios to generate stable long-term cash flows. Although long-term average returns from stocks may remain attractive, a sequence of negative returns can deplete a retirement income portfolio to the point where it lacks sufficient dollars to recover. Retirement assets not only must support a growing population of long-lived investors, but the traditional investment-oriented methods of financing long-term cash flows appear, in the opinion of some commentators, to offer a tenuous prospect for satisfactory results.

One actuarial solution to providing income during retirement takes the form of holding life insurance contracts for the benefit of a surviving spouse and dependents. The demand to hold life insurance contracts during retirement, although an important financial planning topic, is only briefly discussed in this literature review. Rather, most attention is given to the demand to hold annuity contracts. In general, the literature discusses annuities in one of three contexts:

1. As actuarial alternatives to maintaining a portfolio exclusively funded with financial or real assets. Assessment of the annuity approach as an alternative method for producing retirement income often occurs either within a Life-Cycle model context where the goal (objective function)

---

is to maximize the utility of consumption throughout the planning horizon; or, within a shortfall probability context where the annuity is deemed to represent a fail-safe instrument to generate threshold periodic income.

2. As benchmarks for assessing the historical performance of the retirement income portfolio; or, as cost-of-retirement benchmarks for determining the current portfolio’s ability to support its income, gift and bequest objectives.

3. As contracts available within a marketplace of investments that compete for the investor’s dollar alongside of mutual funds, exchange traded funds, real assets, etc. Assessment of annuity solutions, in this context, requires knowledge of contract costs and provisions so that an investor can intelligently compare an annuity to other investment vehicles.

**Historical Back Testing: A Reliable Way to Model the Likelihood of Portfolio Depletion?**

Perhaps the simplest approach for assessing a portfolio’s ability to sustain retirement income is historical back testing (also known as rolling period analysis or overlapping period analysis). As the name implies, the historical return approach calculates the actual returns an investor’s portfolio would have experienced, given its asset allocation. Many retirement income risk models estimate the likelihood of portfolio depletion by specifying a fixed distribution strategy throughout the planning horizon. Furthermore, the static distribution policy itself often operates over a fixed 20, 25 or 30 years. A commonly evaluated strategy is the 4% rule which distributes an annual inflation-adjusted amount equal to 4% of the portfolio’s initial dollar value throughout retirement.³

The historical back testing method tests the success or failure of the retirement plan for each unique planning horizon in the data set of historical returns. The number of unique periods is determined by rolling up the start date of each planning horizon by a single increment of time. For example, L. Bierwirth’s 1994 study, often cited in articles offering financial planning advice,⁴ begins the analysis in 1926 and uses a one year rolling window to calculate 42 unique 27-year rolling periods, ending in 1992. Each sample period is unique by virtue of the fact that its start year drops out of the data set as a new ending year enters the data set. Intervening years, however, continue to appear in multiple samples. Assuming that the past is indicative of the future, the historical model calculates the likelihood of retirement income sustainability by dividing the number of successful planning periods by the total

³ Each year, withdraw four percent of a portfolio’s initial value with an adjustment for annual inflation. The result is a constant-dollar fixed retirement income stream. In essence, the investor self-annuitizes by creating a constant-dollar periodic cash-flow income stream funded for as long as sufficient portfolio assets remain.

number of rolling periods for any given asset allocation / retirement spending strategy combination. The combination with the highest success rate is considered optimal when measured by the likelihood that the unmodified or ‘autopilot’ distribution strategy is sustainable over the applicable horizon. The acceptable retirement income sustainability rate is highly subjective, and depends on investor circumstances and risk tolerance.\(^5\)

Historical back testing is easy to understand, and is a simple way to calculate relatively accurate assessments of what would have happened. However, an investor relying on such a methodology should proceed with caution. This risk modeling approach demands that an investor have faith in the highly dubious assumption that future returns will mimic realized past returns.\(^6\) Although interesting, the pure history model fails to provide assurance that past conditions are sufficiently similar to current conditions so that they act as conditions precedent.

**Sustaining the Portfolio: Static Rules vs. Dynamic Portfolio Management**

Some commentators favor following pre-set, bright-line rules adopted at the time of portfolio implementation. In practice, such rules are sometimes codified within a written Investment Policy Statement. Under this static or ‘architectural’ view of Investment Policy, sticking to pre-set asset management and withdrawal rules offers a high probability of achieving a safe and sustainable lifetime income. Practitioners, however, have recently started to implement more dynamic ‘systems engineering’ approaches to Investment Policy Statements. This shift in emphasis, perhaps arising from severe equity market downturns during the first decade of this century, augments the importance of developing effective methods to monitor the ongoing financial health of the retirement portfolio. Investors and advisors are becoming increasingly uncomfortable with portfolio management based on ‘rule of thumb’ conventional wisdom. Paradoxically, however, some commentators respond to investment turbulence by replacing a single bright-line rule like the four-percent-withdrawal-rate rule with a veritable plethora of bright-line rules-for-all-occasions. This literature review, however, only briefly touches upon the myriad

\(^5\) See DiCarlo, Donald P. and Fast, Steven M. “Prudence: What Are the Odds?” *ALI-ABA Course of Study: Representing Estate and Trust Beneficiaries and Fiduciaries*, San Francisco, California (2008), pp. 477 - 479 for a survey of opinions found in financial advice literature regarding the acceptability of various levels of portfolio shortfall risk. The essay focuses primarily on standards of prudence for management of trust-owned investment portfolios.

\(^6\) McGoun, Elton G., “The History of Risk Measurement,” *Critical Perspectives on Accounting,* Vol. 6, No. 6 (December, 1995): 511 – 532 argues that the empirical distribution of financial asset price returns is not a measure of risk. It is merely a measure of historical realizations which may or may not be applicable to the current economic situation. McGoun’s essay presents a valuable history of risk measurement by economists.
of suggested spending rules. Reliance on a single historical path of realized returns to develop and codify rules for asset allocation and distribution policy is, at the limit, an elaborate exercise in data mining.\textsuperscript{7}

More often than not, an investor’s retirement income preferences are sufficiently complex to call into question the value of standard modeling techniques which implement autopilot formulae like the four-percent-withdrawal-rate rule. Rules-based portfolio management protocols often put the cart before the horse. Most always, actual retirement spending is governed by what the portfolio owner requires to fund the needs of the moment. Testing for portfolio failure rates or for the ‘best’ asset allocation in the context of a monolithic distribution scheme is suboptimal if the objective is to develop financial planning recommendations. For normative research, the tools and techniques of retirement income portfolio evaluation should be sufficiently flexible to accommodate the inevitable changes in client goals and circumstances with which advisors and trustees must grapple.

The literature review demonstrates an evolution away from relatively simple retirement income risk models which apply pre-set distribution rules to static asset allocations towards models capable of capturing realistic and complicated dynamics of investor preferences and behaviors, economic shocks from various exogenous sources, and more credible trajectories of asset returns and inflation rates over a stochastic planning horizon. A credible assessment of portfolio sustainability requires ongoing monitoring of the resources available to discharge future cash flow liabilities.

Dynamic retirement income portfolio management encompasses a variety of strategies. Dynamic programming,\textsuperscript{8} for example, focuses on finding the optimal decision rules given well-defined investor

\textsuperscript{7} One type of data mining occurs when (1) a time-series analysis uncovers patterns or parameters that best fit the sample data; and, (2) the same data is then used to develop and test the efficacy of asset management rules based on the patterns or parameters.

\textsuperscript{8} Dynamic programing is a mathematical theory of optimal sequential decisions under uncertainty. The Bellman Principle of Optimality governs the optimization process: Given an initial state and an initial strategy, the optimal strategy for the next period is the one that would be chosen if the analysis were to begin in the next period—i.e. the strategy for the next period must be optimal given the period two set of conditions arising from the both the decisions made in the initial period and the value of period two state variable(s). The solution over a multi-period horizon is a sequence of single-period optimization problems. The method examines various solution paths (control variables) until the end of the horizon; identifies the best decision making ‘rule(s)’ (the maximization or minimization of an objective function that is a mathematical expression of the investor’s goal); and then traces the optimal path back to the beginning (determines the decision rule(s) for the control variables in each state by backwards induction or other methods). The goal is to find a set of decision making rules that are optimal irrespective of the specific conditions encountered in any single state—i.e. exogenous shocks to state variables do not change the decision rule(s) for control variables. Many Life-Cycle models seek to optimize investor utility (welfare) in terms of a spending control variable. Other commonly found control variables include asset allocation and retirement election (e.g., Social Security or pension benefit) election date. A variation of dynamic programming—optimal control—is commonly used to solve continuous time problems. Boundary control problems attempt to provide optimal decision rules given initial, ongoing, or terminal constraints (boundary conditions) on the objective function—i.e., at least $x$ of wealth must be available at the end of the stochastic planning horizon.
preferences and constraints. Fixed initial conditions (current resources) and well-specified control
variables (jointly considering asset allocation and consumption strategies) often yield the Merton
Optimum in continuous time finance models. The Life-Cycle model literature is an especially fruitful
source for exploring how investors may optimize consumption and bequest objectives in the face of both
deterministic and stochastic variables. However, the investor should recognize the limitation of any rules-
based system derived from dynamic programing models. Different sequences of return and inflation
evolutions, liquidity demands, and other exogenous shocks may produce trajectories manifesting
unacceptable results even when the optimal decision rules are followed. This argues for simulation-based
models that can solve path-dependent problems in which investor utility is both a function of the
portfolio’s value over time as well as the particular path taken to arrive at the value. Dynamic
programming may not be appropriate when the investor faces multiple sources of complexity or sequences
of investing and spending decisions where changes, over time in consumption and wealth values create
complex feedback loops influencing investor risk tolerance.

Additionally, we find a variety of asset management strategies in the literature. For example, a Portfolio
Insurance strategy makes discrete time changes in asset weightings in response to fluctuating portfolio
values in order to replicate a convex option-like payoff structure. Investors may also consider both Buy-
and-Hold—a static approach—and Constant-Mix—a dynamic rebalancing approach—asset management
strategies. Finally, we find a variety of empirical rules [“adaptive portfolio withdrawal strategies”]
purporting to yield maximum likelihood for portfolio sustainability. These rules are often recommended
to investors because they would have produced successful results under previous market conditions. This
is, of course, a variation on the historical back testing method for modeling retirement income portfolios.
The key observation is that the above-listed portfolio management approaches require:

1. Ongoing monitoring of the financial condition of the retirement income portfolio;
2. Careful consideration of the consequences of changes in wealth, distribution requirements, and
   investor circumstances;
3. Identification and intelligent assessment of available asset management planning options; and,
4. An ability to quantify and articulate the probable economic consequences of implementing the
   specific investment options under consideration.

No investment risk model or asset management approach provides a magic bullet for prudent portfolio
management. Both longevity and portfolio sustainability models, and rules-of-thumb based on historical
happenstance, are tools to produce financial insights—they are not substitutes for judgment.
Risk Measures, Benchmarks, and Portfolio Preferencing Metrics

The literature survey presents a number of risk metrics that serve as portfolio preferencing criteria. These metrics include performance evaluation metrics—e.g., the Sharpe Ratio, the Information Ratio, and the Jensen’s Alpha measure; as well as shortfall risk metrics—e.g., shortfall probability relative to periodic income or terminal wealth targets, mean expected shortfall, shortfall magnitude and, in the worst-case, portfolio depletion (bankruptcy) risk or, to use an alternate term, the “risk of ruin.” Performance evaluation metrics help investors make inferences regarding past results. Shortfall risk metrics are forward looking and generate probabilities based either on (1) the assumption that history repeats, or (2) projections derived from an investment risk model. Early research sometimes presents unrealistic outputs from models incorporating questionable assumptions regarding the nature of asset price evolutions and the inflation process. By contrast, more recent research offers a better chance of finding portfolio evaluation methods that include a broad range of variables operating within models producing returns generated under more realistic processes.

The literature survey also evidences a shift in the nature and use of benchmarks. Early empirical studies—e.g., historical back tests—use historical data as a type of implied benchmark against which to judge the prudence and suitability of asset management elections for managing a financial asset portfolio. However, contemporaneously, numerous articles on the topic of annuities appear in actuarial journals. Of particular interest for this literature review is (1) the serious examination of the risks of financing retirement income through ‘self-annuitization’—i.e., maintaining a financial asset portfolio—versus the option to transfer risk to insurance carriers; and (2) the optimality of an annuity-produced (and annuity-constrained) income stream. As financial economists became more familiar with actuarial research, the literature reflects an increased use of an annuity as a comparative benchmark.

Recent research moves beyond historical back-testing and model-based shortfall projections to portfolio assessments based, in large part, on current observables: for example, the cost of lifetime income as embedded in the current price of an annuity. In turn, we see a new set of portfolio monitoring and evaluation metrics focused on retirement ‘feasibility’ and portfolio ‘solvency.’ Simplistically, if retirement assets earmarked for funding lifetime income are worth less than the current cost of an annuity, you can hope that the distribution of realized future returns is sufficiently favorable to overcome the current deficit. Hoping, however, puts a premium on ongoing performance monitoring against relevant benchmarks lest the investor’s hope leads the portfolio towards economic catastrophe.

---

The legal community has more than a passing interest in this discussion. Many irrevocable family trusts contain provisions directing the trustee to provide adequate lifetime income to the current beneficiary and to distribute terminal wealth to remainder beneficiaries. Trust language creates a dual set of claims against trust assets whenever the governing instrument directs the trustee to preserve either the nominal value or the inflation-adjusted value of initial assets for the remaindermen. The current beneficiary holds an income-based claim; the remaindermen hold a claim against terminal wealth. In such cases, trustees must impartially balance the competing claims of each beneficiary class. A failure to fulfill trustee duties may lead to allegations of a fiduciary breach which, if upheld by a court, results in the payment of economic damages from the trustee’s personal assets. Just as the private investor needs something more than hope, the prudent trustee must rely on something more than good faith in the administration of the trust.

The stakes can be high in fiduciary breach litigation; and, therefore, the trust and estate section of the bar began, in the early-1990s, an extensive discussion of investment issues, trust distribution strategies, and portfolio sustainability over the current beneficiary’s lifetime. An important impetus for the discussion was the 1994 adoption of the Uniform Prudent Investor Act by the National Conference of Commissioners on Uniform State Laws. Usually, this discussion is confined to legal journals and monographs most of which fail to appear as bibliographical references in investment- and actuarial-oriented journals.10

**Utility Maximization and Shortfall Minimization**

Generally, one finds two types of retirement income models. Life-Cycle models11 encompass a range of models in which investors utilize information to make sequential decisions to facilitate the attainment of financial goals. Although not all Life-Cycle models are utility (welfare) maximizing models, most assume that the investor attempts to smooth lifetime spending—i.e., maintain a constant marginal utility of consumption. Stochastic dynamic programming is a popular method for identifying asset management strategies producing the greatest aggregate utility over the investor’s lifetime.12

---


12 When variables evolve continuously—a diffusion—they are modeled as a continuous-time stochastic process.
By contrast, shortfall minimization models often test autopilot asset allocations and spending rules. Historical back tests, bootstrapped reshuffling of historical returns, and Monte Carlo simulations of pre-parameterized distributions are popular methods of projecting the likelihood of a shortfall should a myopic investor elect to stay the course throughout all economic conditions.\(^\text{13}\)

In some cases, a shortfall minimization preferencing criterion leads to allocations and withdrawal strategies that differ substantially from those recommended by Life-Cycle models purporting to optimize investor utility. An important research advance occurred in the mid-2000s as several articles consider whether it is possible to reconcile the two types of retirement income risk models. Generally speaking, a condition for resolution occurs when:

1. The utility function being optimized in Life-Cycle models shifts away from Constant Relative Risk Aversion—risk aversion is invariant to changes in the level of wealth—to either Hyperbolic Absolute Risk Aversion—risk aversion (decreasing or increasing) is a possibly linear function of the wealth level—or, to a form of state-preference utility in which asset payoffs in one economic (or, health) state are preferred to equivalent or greater payoffs in other states. Aggregate utility calculated by applying a homogenous utility function over every economic state—i.e. across a distribution of possible investment results—gives way to subjective preferences that differ across each state—i.e., state-contingent preferences.

2. A minimum threshold level of income or wealth is introduced into the model. The presence of such a consumption floor means that the investor derives utility only for amounts equal to or in excess of the target income or wealth level.

The importance of reconciling preferencing criteria for minimizing shortfall risk and for enhancing utility of consumption and terminal wealth cannot be overstated. This line of research also motivates (1) a reassessment of the role of annuities as a wealth-management tool; and, (2) using an annuity-based benchmark to compare investment strategies. Furthermore, it provides intellectual underpinning for assessing the financial health of a retirement income portfolio in terms of solvency and feasibility metrics; and it opens portfolio monitoring to a richer set of risk metrics extending well beyond the likelihood-of-failure projections. This is the subject of the next section.

\(^{13}\) This set of metrics includes a variety of risk measures including the likelihood of a shortfall in either periodic consumption or terminal wealth; an unacceptable variance in periodic consumption; the expected value of a shortfall either in terms of the Present Value of a shortfall or in terms of the mean expected shortfall given that a shortfall occurs; the magnitude or duration (time without funds) of a shortfall, a distribution of shortfall results, etc. A research monograph sponsored by the Society of Actuaries’ Pension Section and Pension Section Research Committee [Measures of Retirement Benefit Adequacy: Which, Why, for Whom, and How Much?" Bajtelsmit, Vickie, Rappaport, Anna and Foster, LeAndra (2013)] provides a helpful review of shortfall risk metrics. (https://www soa org).
Assessing the Retirement Income Portfolio: Solvency v. Shortfall

Sustainability of adequate lifetime income is a critical portfolio objective for retired investors. Commentators often define sustainability in terms of (1) a portfolio’s ability to continue to make distributions throughout the applicable planning horizon, or (2) a portfolio’s ability to fund a minimum level of target income at every interval during the planning horizon. The first approach focuses on the likelihood of ending with positive wealth, or, if wealth is depleted prior to the end of the planning horizon, on the magnitude and duration of the shortfall; the second focuses on the likelihood of consistently meeting all period-by-period minimum cash flow requirements.

Risk models help advisors assess a portfolio’s ability to provide adequate cash flow throughout retirement. Conclusions about cash flow sustainability are usually reached by determining the likelihood that distributions (fixed amounts, percentage of corpus, or ‘dynamic’) can be maintained for either deterministic or stochastic time periods under various asset allocations and longevity assumptions. It is the risk model that generates the distribution of future results; and, therefore, probability assessments are not independent of the model.14

A critical distinction must be drawn between sustainability—the probability calculated by the risk model that future financial market returns are sufficient to defease targeted cash flows, and feasibility—a judgment regarding the current ability of the portfolio to fund the present value of required cash flows. The feasibility condition requires that the current market value of assets equals the stochastic present value of the lifetime target income plus, if relevant, gifts and bequests. Conceptually, it is useful to consider portfolio monitoring as a test of a portfolio’s funded status—i.e., as an activity occurring within the context of ‘free boundary’ problems.

One class of free boundary problems, known as Stefan problems, involves estimating the demarcation between two regions where the line of demarcation is not fixed.15 A classic example is estimating the

---

14 Albrecht, Peter, Maurer, Raimond and Ruckpaul, Ulla, “Shortfall-Risks of Stocks in the Long Run,” *Financial Markets and Portfolio Management* Vol. 15, No. 4 (2001): 481 – 499, provide a detailed assessment of shortfall risk measures. They conclude: “…the use of the shortfall probability alone is insufficient for the assessment of the risk of stock investments in the long run….the probability of a loss or a shortfall decreases with the length of the time horizon. However, the average level of the loss or the shortfall respectively, given a loss or a shortfall has occurred, increases.” [P. 488]

15 Friedman, Avner, “Free Boundary Problems in Science and Technology,” Notices of the American Mathematical Society (September, 2000): 854. Free boundary problems require solutions to differential equations. An introduction to these equations within the context of trust portfolio management is provided by Collins, Patrick J, and Stampfli, Josh, “Promises and Pitfalls of Total Return Trusts,” *ACTEC Journal* Vol. 27 (Winter, 2001): 205 - 219. A partial differential equation is an equation that reflects the simultaneous rates of change of several independent variables. Not all partial differential equations have a closed-form (analytical) solution. When risk modeling demands (1) considering various decision points over time, (2) incorporating uncertainty from various...
location of the boundary between solid ice and liquid water when the temperature drops below freezing. In winter, the depth of a Minnesota lake’s boundary between ice and water fluctuates according to the random variable of water temperature. In cold weather, the ice pack is thick; in warmer weather, it thins. Analogizing to assessing a retirement income portfolio in terms of the feasibility condition, the investor faces the problem of determining the line of demarcation between two regions—a region of wealth surplus, in which the portfolio’s current value is able to support financial objectives; and a region of wealth deficit, in which the portfolio’s current value is not able to support financial objectives. A bull market tends to move a portfolio farther into the feasible region—the region of wealth surplus. A bear market, however, tends to move it toward the line of demarcation—the free boundary—that separates the regions. As wealth depletion pushes farther and farther toward the region of infeasibility, the consequences of investor actions or inactions are magnified in the sense that an ill-considered asset management election may not only generate losses, but losses from which the portfolio can never recover. The investor has to know, in effect, the thickness of the ice pack, lest undue optimism induces the fatal mistake of conducting financial affairs on thin ice.

The literature review presents the intellectual background for considering management of a retirement income portfolio as a free boundary problem. In many respects, the research history expresses the retirement planning solution in terms of a utility based model in which the boundary is defined either in option valuation terms—e.g., optimal time to annuitize—or as a boundary control problem in which wealth is not allowed to drop below either the stochastic present value of liabilities, a periodic consumption level, a utility value, or a similarly defined demarcation level. The contribution of legal commentary to this discussion is its focus on the economic consequences of portfolio management elections in terms of the legitimate economic expectations of interested parties where such expectations are couched in terms of terminal dollar wealth and ongoing income levels. The contribution of the actuarial literature is the analysis of annuity solutions to the portfolio sustainability problem. The insights derived from this branch of intellectual study enables researchers to redefine the retirement income problem in terms of annuity pricing formulae that calculate the present-day cost for providing future periodic income.

This literature review uses the term “free boundary” in several contexts: (1) as an element in a barrier control problem, (2) as an accounting concept—current value of assets minus stochastic present value of liabilities, (3) as a benchmark-relative measure—i.e., an annuity cost benchmark, (4) as a solution point for a differential equation in the classic mathematical definition of free boundary problems, and (5) as a sources, or (3) using complex systems of independent variables, numerical methods are required to approximate a solution.
reference point for both prior research and ongoing research in the field of retirement income portfolio monitoring and evaluation. Benchmarking the retirement income portfolio in terms of an objective, market-determined cost benchmark enables researchers to create models that define a ‘free boundary’ as the point at which a portfolio becomes technically insolvent because it violates the feasibility condition—assets are less than liabilities. Specifically, absent bequest objectives, the free boundary lies at the point where the current market value of assets is less than the current cost of a single premium annuity priced to provide a minimum acceptable lifetime income. Several recent retirement income models express longevity risk in terms of annuity cost. Longevity risk reduces to a dollar value!

Monitoring the feasibility condition differs from assessing a portfolio’s financial health in terms of the likelihood of sustainability over the planning horizon. As noted, the likelihood of future portfolio depletion often expresses a shortfall probability derived from projections made by risk models employing Monte Carlo simulation or bootstrap analysis. Unlike the feasibility condition which is a function of current market observables, the shortfall probability metric is the risk model’s best guess. Depending on the projected time of portfolio depletion, the risk to the investor may be severe if other assets—e.g., social security, home equity, family support—cannot cushion the magnitude and duration of the shortfall. Advances in modeling however facilitate prudent portfolio surveillance and monitoring by tracking both shortfall and feasibility risk metrics. Over time, the locus of action in retirement income planning shifts away from a search for the best portfolio design and spending rules to dynamic portfolio monitoring and intelligent assessment of asset management elections.
PART ONE: RETIREMENT INCOME RISK MODELS [1965 – 2005]

Setting the Stage: Optimizing Retirement Consumption / Minimizing Shortfall Probability

Many academic commentaries trace their origin to the 1965 paper by Menahem E. Yaari [“Uncertain Lifetime, Life Insurance, and the Theory of the Consumer”] which demonstrates that investors lacking a bequest objective and having access to actuarially fair annuities in a complete market setting—i.e., insurance or financial instruments span all economic risks faced by the investor—will hold all wealth in “actuarial notes.” [P. 140] Yaari’s study considers how the investor can optimize discounted expected utility over both a deterministic period and an uncertain life span where the investor places a value on both consumption and bequests. Investors lacking a bequest objective will maximize a utility function for consumption only; and their challenge is to find the optimal feasible consumption plan when the planning horizon is uncertain. Under the Yaari complete market model, annuities put the investor on the optimal feasible consumption path.

When a bequest motive exists, however, the investor must solve for optimization of two decision variables: a feasible consumption plan and a feasible saving plan. Assuming no labor income, consumption is funded entirely with annuities, while the saving plan’s asset allocation is a function of available investment returns. In equilibrium, the marginal utility of the consumption plan will exactly equal that of the saving plan. Yaari makes the important observation that when annuities are available, “…the consumer can separate the consumption decision from the bequest decision.” [P. 149] In the absence of annuities, such a separation is not possible. As this literature survey unfolds, it points out how future studies build on Yaari’s observations whenever their authors recommend a “two-fund” solution (a guaranteed annuity income plus investment of surplus wealth in a performance-seeking portfolio) to the retirement income/bequest challenge.

Yaari’s paper is also an important source document for studies applying a portfolio shortfall probability risk metric: “The chance-constrained programming approach requires that the constraint (in this case the wealth constraint) be met with probability \( \lambda \) or more, where \( \lambda \) is some number fixed in advance, say .95…” [P. 139] Under this evaluation metric, the optimal portfolio minimizes shortfall probability. However, Yaari does not develop a probability-of-success approach to evaluating portfolio choice. Rather, he evaluates investment and actuarial solutions in terms of maximizing investor utility (welfare) over the applicable planning horizon at an appropriate discount rate for a risk-averse investor manifesting a possibly time-varying preference for consumption as a function of age. Yaari’s model sets the table for future inquiry in that it acknowledges (1) the usefulness of both utility-based, life-cycle modelling and shortfall risk modelling; (2) the existence of both actuarial and investment oriented solutions; and (3) the
importance of the consumer’s preference function ("impatience") with respect to the timing and magnitude of retirement spending.

By 1997 Milevsky, Ho & Robinson’s study [“Asset Allocation Via the Conditional First Time Exit or How To Avoid Outliving Your Money”] highlights the significance of shortfall probability as a preferencing criterion for retirement income investment strategy. They phrase the optimization problem in terms of projecting the time at which a portfolio, maintaining a fixed vector of asset allocation weights and earning a stochastic return, suffers depletion under the stress of spending requirements. The optimal asset allocation postpones the time of depletion to the furthest feasible future date. The paper decomposes wealth into asset and liability components: net portfolio value at any future date consists of the compound growth of initial wealth less the income payout liability defined as the accumulated value of an annuity due. The date of ruin—the conditional first time exit—is the limit of time for which wealth is greater than consumption demands. Beyond this time, the investor will run out of money. Additionally, the paper is one of the first to argue the need for dynamic asset allocation in the face of a static spending policy—especially where there is a lower bound designating an acceptable spending level. The study is an early example of valuing retirement liabilities by using annuity pricing factors. Although the authors are primarily interested in solving for the date of financial ruin, their approach is generally conformable to a free boundary approach. The existence of a threshold, or minimum, periodic income need throughout each retirement period becomes an increasingly important aspect of risk modelling in future studies.

Milevsky published a follow on study in 1998 [“Optimal Asset Allocation Towards the End of the Life Cycle: To Annuitize or Not to Annuitize”] which presents and explores the importance of the Wealth/Consumption Ratio in identifying the optimal asset allocation. The paper introduces the topic of the optimal time to exercise an option to annuitize financial wealth. Milevsky expresses the optimal time to annuitize in option valuation terms. A threshold condition for annuitization occurs when annuity mortality credits overtake annuity costs. The recommended course is to evaluate the portfolio after each period. If there is surplus wealth—i.e., the portfolio’s current market value is greater than the present value of projected future spending—then it is beneficial to delay exercise of the annuitization option. A primary justification for delay is that early exercise destroys the option’s time value—all else equal, the cost of an annuity-provided income stream should decrease as the investor ages. Delay, however, is not risk free because the annuity’s future costs may rise as a function of interest rate changes. The delay-for-as-long-as-possible strategy, however, implies that prudent portfolio management does not allow financial wealth to drop below the present value of consumption. An individual is better off deferring purchase of
an annuity if his after consumption wealth at the beginning of period \( x \) is greater than the annuity’s cost at the beginning of period \( x+1 \).\(^{16}\)

Whereas the purchase of an annuity is an irrevocable decision to pay a non-refundable sum to an insurance company, and whereas the decision eliminates liquidity and the ability to make bequests, the investor should defer annuitization, according to Milevsky, for as long as possible provided that the risk of failing to acquire an adequate lifetime income stream remains within tolerable levels. Later studies by Milevsky and others [e.g., the 2002 essay “Optimal Asset Allocation and The Real Option to Delay Annuitization: It’s Not Now-or-Never,” co-authored with Virginia R. Young], de-emphasize the importance of the Wealth/Consumption Ratio in favor of a pure option-to-annuitize valuation approach. They argue that holding investments in a risky asset portfolio beyond the optimal stopping time could result in a further deterioration in portfolio values to the detriment of the investor. Milevsky’s shifting approaches might, at first read, seem contradictory because they lead to seemingly different conclusions regarding preferred portfolio management elections. However, the apparent contradiction is more the byproduct of differing mathematical explorations of decision making approaches than it is the byproduct of incongruous thinking.

Philip Dybvig’s 1999 paper [“Using Asset Allocation to Protect Spending”] is one of the first essays to explore the linkage between asset allocation policy and portfolio spending policy. Dybvig argues that asset allocation and spending decisions must be made jointly. Common practice is to link spending to the long-term expected return of the strategic asset allocation. This is a static linkage and does not provide dynamic feedback to the investor. Dybvig, like Milevsky, advocates a sequential (year-by-year) decision making process. Asset allocation changes dynamically to reflect changes in the value of wealth, interest rates, and spending objectives. Both Dybvig’s and Milevsky’s papers emphasize the importance of (1) monitoring, (2) periodically evaluating asset management elections, and (3) active portfolio administration. As such, they constitute important source material justifying active management of the retirement income portfolio.\(^{17}\)

**Investment v. Actuarial Solutions**

Several papers published during this period explore the question of whether annuitization, especially in periods of low interest rates, provides either an optimal income stream or a preferred portfolio management strategy. For example, the 1999 study by Mitchell, Poterba, Warshawsky and Brown [“New

\(^{16}\) This begs the question of how an investor projects the annuity’s cost in period \( x+1 \). The distribution of the future cost of annuity contracts assumes increasing importance in future research papers.

\(^{17}\) We distinguish between active portfolio management, which requires ongoing portfolio monitoring and review, and active investing which is the attempt to ‘beat the market.’ Investment texts often contrast active investing to passive investing which uses index funds in an attempt to ‘match the market.’
Evidence on the Money’s Worth of Individual Annuities” notes that annuities are bond-backed portfolios acquired by insurance companies operating in the capital markets. Whereas the duration of an annuity is, in general, greater than the duration of an underlying bond portfolio, an investor who buys a nominal dollar annuity exchanges systematic longevity risk for systematic interest rate risk. The 2001 essay by Peter Albrecht and Raimond Maurer [“Self-Annuitization, Ruin Risk in Retirement and Asset Allocation: The Annuity Benchmark”] concludes that annuities purchased in low interest rate environments produce modest payouts that, in most cases, can be achieved through self-annuitization of a risky asset portfolio. However, at older ages and in higher interest rate environments, attempting to match an annuity payout with distributions from a risky asset portfolio runs a much higher risk of ruin.

These observations have implications for investors in down market economies where interest rates tend to be low. Jeffrey R. Brown, Olivia Mitchell and James Poterba’s 2001 study [“The Role of Real Annuities and Indexed Bonds in Individual Accounts Retirement Program”] points out that maintaining the constant dollar value of consumption may not be the optimal consumption path. Smoothed consumption is the ability to keep the marginal utility—as opposed to the dollar value—of expenditures steady. Constant (real) annuity income may not reflect the time preferences of retired investors. Interestingly, J.R. Brown’s paper also published in 2001 [“Private Pensions, Mortality Risk, and the Decision to Annuitize”] suggests a possible solution. If there is sufficient surplus wealth, the investor may wish to annuitize most wealth so that there is a choice to spend most or all of the periodic payment if the time preference discount rate is high, or to save and reinvest a portion of each payment if the time preference discount rate is low. Brown’s observations suggest that annuitization might be considered even when current wealth is well above the free boundary location. Part Three of this literature review provides additional discussion on this topic.

James M. Poterba presented a paper entitled “Annuity Markets and Retirement Security” at the Third Annual Conference of the Retirement Research Consortium on May 17 – 18, 2001 in Washington DC. Poterba’s presentation links the topic of the optimal retirement income stream with the concept of option-based optimal stopping time for annuitization. Optimal stopping time approaches to asset management decision making—i.e., the optimal time to switch from a risky asset portfolio to annuity-generated income—constitute an important alternative to the free boundary approach. Poterba notes that annuitizing all wealth at a single moment (“an optimal stopping time”) may not be optimal. It is, in his opinion, a type of annuity market timing. Poterba develops the theme of ‘time diversification’ of annuity purchases. Variation in bond returns generates a substantial variation in annuity payout rates over time; and, therefore, annuitizing all wealth at a single moment is a risky strategy.
The literature reveals an ongoing tension among commentators on the subject of optimal annuity timing. Some suggest that investors should annuitize all wealth at the point in time when the annuity contract’s mortality premium exceeds the expected premium from holding a risky-asset portfolio—an option valuation approach that stresses the fact that annuities are a cost-efficient way to generate income. Others suggest that investors should annuitize all wealth as soon as they can lock in their desired future income stream (‘a bird in the hand….’). Finally, others suggest that, if prudent, investors should delay exercising the option to annuitize. The recommendation to delay stems from the observations that (1) immediate annuitization may impose an unacceptable constraint on future consumption; and, (2) investors value smoothed marginal utility of lifetime consumption more than a fixed periodic dollar-value income.

We can illustrate something of the nature of this ongoing debate by jumping ahead to an article written ten years after Poterba’s presentation: Moshe Milevsky and H. Huang [“Spending Retirement On Planet Vulcan: The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates” 2011]. They echo Poterba’s arguments regarding the optimality of securing retirement income by annuitizing all wealth at a single moment. However, they make their case within the context of maximizing a utility of consumption function: “…the utility-maximizing retiree is not willing to reduce their initial standard of living simply because of a small probability they will reach age 105….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 and linked to their risk aversion.” [P. 48] The authors quote Irving Fisher (1930): “…[the investor has] a high degree of impatience for income because 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?” [P. 47] We expand the discussion of the Irving Fisher view of utility maximization in Part Three when we examine more closely the calculation of an annuity’s utility value under Life-Cycle models of consumption. Part Three also charts the impact of more complex Life-Cycle retirement risk models that incorporate other utility functions such as Epstein-Zin, state-preference, and ‘habit’ utility.18

It is interesting to contrast Poterba’s view with that expressed by John Americks and Paul Yakoboski in 2003 [“Reducing Retirement Income Risks: The Role of Annuitization”]: “…there is an inherent tradeoff between maintaining a stock of assets and supporting a flow of income…” [P. 18] The annuity contract,

---

18 Yaari acknowledges the importance of Fisher Utility and includes a term for the investor’s subjective discount rate (impatience) that may differ from the prevailing risk-free rate. Consumption is decreasing “…whenever the rate of subjective discount is greater than the rate of interest.” [P. 138] Yaari, however laments that research models incorporate a number of simplifying assumptions (“…one makes assumptions to suit the problem at hand…” [P. 137]; and that the form of his model’s utility function means “…that the consumer’s preferences are independent over time. This is not a very happy assumption but unfortunately some strong assumption of this sort must be made in order to make the problem manageable.”) [P. 137] It is not too far off the mark to assert that future research studies either seek to address and overcome such limitations, or simply ignore them.
according to Americks and Yakoboski, is valuable in so far as it allows retirees to achieve the greatest efficiency in spending money throughout retirement. Provided that annuity contract loads are reasonable, the authors point out that annuitization produces a dollar of lifetime income for older age investors at a cheaper cost than any other investment alternative. The debate between annuitization and self-annuitization takes center stage during this period. Recommendations concerning the optimal time to exercise an annuitization option reflect either the cost-efficiency school of thought—e.g., mortality credits dominate the expected risky asset portfolio premium—or, the utility maximization school of thought—e.g., a dynamic programming approach to the retirement consumption objective.

J. Michael Orszag, acting either as sole author or as a co-author, contributes several studies in 2002 to the cost-efficiency school of thought. In the paper “Discrete-time Drawdown Analytics: Annuities and Drawdown in a Retirement Income Model,” he takes an option valuation approach in that the optimal time to annuitize is the point at which the annuity produces an income stream higher than a portfolio drawdown program. Thus it is the ratio of consumption from the portfolio to consumption provided by the annuity that is determinative of the optimal time for annuitization—not the wealth-to-consumption ratio. Whereas annuity mortality credits become greater at older ages, Orszag generally recommends a top-down approach in which annuitization is delayed until an optimal stopping time. However, in a nod to the merits of utility-based arguments, he notes that annuities lock in the budget constraint. This means that there is a risk of losing potentially higher income at later ages if annuitization occurs too early in retirement.

In a study co-authored with Sandeep Kapur, [“Portfolio Choice and Retirement Income Solutions”], Orszag further develops the concept of the ‘annuity premium’—the spread between annuity yields and long-term government bond yields. Capturing this spread requires “capital sacrifice.” [P. 2] Thus, a decision to acquire an immediate payout annuity to protect the current standard of living involves defending capital sacrifice. The annuity premium crowds out bonds from the portfolio.19 If, however, equity is used to purchase an annuity, the decision is to exchange the expected equity risk premium for the annuity premium. Kapur and Orszag contend that whenever the annuity premium exceeds the expected equity risk premium, the option to annuitize should be exercised. This is a preference-free calculation—a question of stochastic dominance rather than investor utility or target income feasibility.

---

19 An annuity pays bond interest plus an extra amount to contract holders for as long as they live. Ultimately, the bond defaults and fails to return principal—but this event occurs only at the annuitant’s death. The default is, in fact, the source of the extra amount—the annuity premium—enjoyed by all annuitants participating in the mortality-pooling arrangement.
However, perhaps in an effort to bridge the two approaches to the retirement consumption problem, Orszag presents a third essay [“Ruin in Retirement: Running Out of Money in Drawdown Programs”]. The paper reintroduces a preferencing criterion by establishing a lower-bound consumption target. The consequences of introducing a threshold lower income or wealth bound into the model are profound. In the third paper, Orszag solves a differential equation to determine the income withdrawal rate target that is the reciprocal of the consumption target. The withdrawal rate calculation, however, ignores the influence of stochastic mortality. Orszag defines it as the rate that will drive the portfolio value to zero after ‘x’ number of years, where ‘x’ is a fixed number. Orszag, in contrast to Poterba, argues that annuities are more attractive than bonds in low interest rate environments because the annuity mortality premium becomes a greater percentage of total current income. Orszag’s view also contrasts, for example, with that expressed, in 2003, by Peng Chen and Moshe Milevsky [“Merging Asset Allocation and Longevity Insurance: An Optimal Perspective on Payout Annuities”]. Chen and Milevsky point out that payouts on annuities are a function of the prevailing interest rate at the time the contract is executed. As a consequence, “locking-in a fixed annuity is implicitly a market timing play.” [P. 4]

Gabriele Stabile [“Optimal Timing of the Annuity Purchases: A Combined Stochastic Control and Optimal Stopping Problem,” 2003] defines a region for which it is not optimal to annuitize as well as a region in which the investor immediately annuitizes all wealth. The regions are separated by a boundary where the boundary is defined by a utility-of-wealth value function rather than by a dollar-denominated free boundary. The optimal time for annuitization occurs when the dynamic programming solution to the value function indicates that investor preferences are best served through an actuarial as opposed to an investment-oriented solution. If the investor’s utility-adjusted wealth remains above the lower threshold, then the investment program continues—i.e., annuitization is deferred. Her risk model indicates, if wealth remains greater than a threshold amount, the investor continues to participate in financial markets even in the case where the investor’s risk aversion for income derived in financial markets is higher than for annuity payouts. If the investor approaches the boundary from below, however, the decision rule is to purchase an annuity at the boundary. If the investor does not immediately annuitize at the boundary, the situation might deteriorate further.

Russell Gerard, Steven Haberman and Elena Vigna [“Optimal Investment Choices Post Retirement in a Defined Contribution Pension Scheme,” 2004] provide an interesting variation of the boundary problem and a counterpoint to Stabile’s observations. They point out that the region of annuitization might occur either at a minimum boundary to protect against further deterioration of wealth, or at a maximum boundary which is defined as the fund size sufficient to purchase a large lifetime income stream. If the

---

20 The implications of sustaining a threshold income form a major topic in Part III of this literature survey.
annuitization benefit is sufficiently large, there is little need to continue investing in a risky asset portfolio—the utility of the extra dollar generated by investing grows vanishingly small. The authors’ model incorporates a quadratic loss (disutility) function when the investment portfolio fails to maintain an amount sufficient to purchase a single premium immediate annuity for the investor’s target income. The lower an investor’s risk aversion, the higher the future income target that the financial asset portfolio must fund. Interestingly, a utility penalty is assigned both when the investment portfolio under-performs its target and when it over-performs. Under-performance indicates that the desired standard of living is at risk; over-performance indicates that the portfolio could have generated the required return with less risk. The authors estimate optimal asset management strategies for both a constant income target and an exponentially increasing income target. Given the model’s assumptions regarding quadratic risk aversion, it is not surprising that the optimal solution is the Merton optimum adjusted for the investor’s risk aversion with respect to the income and bequest goals.

The Gerard, Haberman and Vigna article is an early example of a series of works focusing on the cost of retirement. Some commentators argue that, absent a bequest motive, unspent funds represent a missed opportunity to improve lifetime living standards; others argue that overfunding retirement income targets misaligns financial goals (liabilities) with the assets designed to fund them. See, for example, the 2009 publication by Jason Scott, William Sharpe, and John Watson [“The 4% Rule—At What Price?”] which asserts that a truly risk-free retirement consumption strategy never exhibits either a surplus or deficit. Under this view, any strategy that matches constant spending with volatile (uncertain) return outcomes is suboptimal in several respects: (1) there is a positive probability of ruin; and (2) there is a cost to generate an unneeded / low-utility surplus.

Milevsky reappears in our survey of academic literature. Moshe Milevsky, Kristen S. Moore and Virginia Young co-author “Optimal Asset Allocation and Ruin-Minimization Annuitization Strategies: The Fixed Consumption Case,” which employs a barrier control problem approach to the issue. When current wealth is greater than the market price of a target annuity income stream, the investor will annuitize to lock in the targeted lifetime consumption. However, when wealth is insufficient, the question becomes the optimal time at which to buy the annuity: “…the optimal annuity-purchasing scheme is a type of barrier control.” [P. 12] To the left of the barrier—wealth is below the stochastic present value of consumption—the investor makes no annuity purchase; to the right of the barrier, the investor will buy an annuity sufficient to guarantee the targeted periodic income.
**Moving beyond the Constant Relative Risk Aversion Assumption: A Minimum Income Need**

In 2003, Vanduffel, Dhaene, Goovaerts, and Kaas explore the topic of how much money (reserves) an insurance company needs in order to assure funding adequacy for a stream of future liability payments at a given confidence level [“The hurdle-race problem”]. Although the article’s primary intended audience is insurance company actuaries, the authors develop a portfolio management approach that forms a solid basis for ascertaining portfolio sustainability and for implementing procedures to protect the portfolio’s ability to provide required future payments. If one wishes to meet payment obligations with certainty, the replicating portfolio consists of ‘n’ zero-coupon bonds assuming that the liability is deterministic. The study, however, also calculates the optimal reserve when the investment portfolio generates stochastic rather than certain returns. In this case, the objective is to determine the reserve or, provision “…such that the probability that we will be able to meet our future obligations will be sufficiently large. Conversely, if the level of the provision is given, our methodology will enable us to compute the probability that we will be able to meet our future obligations under the given investment strategy.” [P. 405] When the level of the provision or reserve is given—such as the case for a retirement income portfolio—“…the optimal investment strategy could be determined as the one leading to the maximal probability that we will be able to meet our future obligations.” [P. 406]21

The Hurdle-Race Problem paper details how to calculate the provision when the future payments are known and when the provision is invested in a stochastic return process. Given a stream of liability payments, the provision must have a value equal to or greater than zero at the end of the applicable planning horizon. The reserve is adequate if its level is greater than the stochastic present value [PV] of the payments to be made. The liability is deterministic but the present value of the liability payment stream is subject to changes in the discount rate. Therefore the PV of the liability is also stochastic. A limitation of defining the optimal reserve in terms of “reaching the finish”[P. 407] is that there may be situations where the interim value of the reserve falls below a threshold level which may violate regulatory requirements or which may, in general, represent an undesirable situation. The problem becomes one of calculating the optimal reserve not only in terms of the ultimate goal but also in terms of the period-to-period reserve value: “…the conditions that year-to-year the provision $R_j$ is larger than a given deterministic value $V_j$ with a sufficiently large probability. These additional requirements are the ‘hurdles’ that have to be taken.” [P. 407] The authors’ approach shares many similarities with a free

---

21 The reader will recognize this as a variation on the “chance-constrained programming approach” introduced—but not developed—in Yarri’s 1965 study. A more recent variation on the “hurdle race” approach to asset management is the 2008 essay by Jason S. Scott [“The Longevity Annuity: An Annuity for Everyone?”] in which Scott suggests that the initial “provision” might be relatively small if the investor is willing to purchase a contingent payout life annuity that provides a periodic income only to contract holders who survive to an advanced age.
boundary framework for prudent asset management particularly with respect to meeting interim “hurdles” or solvency checkpoints as part of monitoring the retirement feasibility condition.

We can draw a direct line from Yaari’s 1965 article to the 2003 study by David Blake, Andrew Cairns and Kevin Dowd [“Pensionmetrics 2: stochastic pension plan design during the distribution phase”]. Yaari, as noted, opts to use a utility-of-consumption approach for determining the optimal consumption path for investors with limited resources and no bequest objective. However, in order to achieve mathematical tractability, Yaari, as well as most commentators following in his footsteps, assumes that the risk aversion function remains constant despite changes in the level of wealth. Although it had long been recognized that constant relative risk aversion applies to investors only under a limited number of conditions, nevertheless, through the end of the 20th century, many studies evaluate mathematical models based on the twin assumptions of log normal risky asset return distributions (geometric Brownian evolutions) and constant relative risk aversion on the part of investors. These assumptions have profound consequences. If return distributions are similar to the bell curve, then a sequence of below-the-mean returns is merely a series of unlucky draws from a stable distribution. Realized results carry little information regarding the desirability of holding the risky asset portfolio in future periods, and changes in investment strategy may be unwarranted because the nature of the underlying distribution—i.e., the probability of future sequences of positive or negative returns—remains unchanged. Likewise, the assumption of constant relative risk aversion assumes that an investor maintains full willingness to continue a risky venture despite the pressure on current surplus—resources above the free boundary line. Operating together, these twin assumptions provide a powerful rationale for staying the course, trusting in the restorative power of capital markets, and maintaining equanimity in the face of decreasing wealth.

Blake, Cairns and Dowd, by contrast, restore the focus to the wealth/consumption ratio. They argue that optimal annuitization is a function of investment performance and the size of the wealth fund. Smaller sized accounts produce less income and the marginal utility of the excess income from annuitization has greater value: “…the marginal utility of consumption gets large as the fund size gets small.” [P. 43] The annuitization trigger changes from an optimal stopping time problem—annuitize when the annuity mortality premium exceeds the expected equity risk premium—to a decision directly tied to portfolio size relative to the distributional demands: “…a larger fund size makes it more likely that the plan member will delay annuitization.” [P. 45] This important study is one of the first to model several different utility functions to ascertain how output changes with changes in the utility function’s form.

---

We also refer readers to an excellent 2003 discussion [“Annuities and Individual Welfare” Thomas Davidoff, Jeffrey Brown and Peter Diamond] regarding the economic assumption of complete markets—an assumption that starts with Yaari and continues throughout the history of academic commentary. Davidoff, Brown and Diamond relax Yaari’s complete market assumption and demonstrate that, absent a bequest motive, investors will annuitize some wealth as long as annuities pay a higher return than assets of comparable risk.

**Utility Maximization and Shortfall Probability Minimization**

By 2005, several research paths find an elegant synthesis in the article “Normative Target-Based Decision Making” by Ali E. Abbas and James E. Matheson. The importance of this paper rests, in part, in its observation that utility-based decision making is compatible with shortfall or target-based decision theory in the presence of a required minimum level of income that must be maintained throughout the applicable planning horizon.

Classical utility theory states that investors seek to optimize utility over the distribution of potential outcomes. However, the presence of a lower bound creates a type of “step-utility function” [P. 373] which divides outcomes into two regions—acceptable and unacceptable. With suitable mathematical transforms, the utility of the investment project is equal to the probability that the result is above an “aspiration level” [P. 377] which divides the two regions. Although the aspiration level is the boundary between satisfactory and unsatisfactory, its location constantly changes because of “changes in the lottery that the individual is facing.”[P. 377] For investment issues, the aspiration level changes with factors such as wealth level and liability values. Probability distributions may, however, change through time and the targets may have to be revised to reflect updated information: “…pursuing a fixed goal may be operationally motivational when things are going smoothly, but when major impacts, such as setbacks or new opportunities, create a need to re-evaluate alternatives, the normative approach demands determining new targets.… Simply maximizing the probability of reaching the old target is no longer optimal.” [P. 384] This paper provides solid underpinning for asserting the importance of active monitoring and portfolio surveillance in terms of the investor’s retirement income goals. We also note the 2003 paper by Michael Stutzer [“Asset Allocation Advice: Reconciling Expected Utility and Shortfall Risk”] which

---

23 The utility function, at the point where it reaches the ‘aspiration level,’ is found to exhibit a discontinuity in other research. See, for example, Diecidue, E. and van de Ven, J. 2008. “Aspiration level, probability of success and failure, and expected utility.” *International Economic Review*, Vol. 49, No. 2: 683 – 700.
discusses differences between the classical approach of portfolio optimization based on utility maximization vs. approaches based on shortfall probability minimization.²⁴

Both the Abbas & Matheson and the Stutzer papers argue that the presence of a consumption floor—a lower boundary based on a required periodic income stream—changes the nature of the decision making process. A 2005 paper by Geoffrey Kingston and Susan Thorp [“Annuitization and asset allocation with HARA utility”] further advances this argument and, in doing so, clarifies the relationship between investor utility (welfare) and portfolio management elections. The article advances a rationale for a ‘bottom-up’ approach to annuitization—i.e., annuitize-as-soon-as-possible—similar to the one made several years later by David F. Babbel and Craig B. Merrill in their 2007 paper “Rational Decumulation.” The minimum standard of living target is fully funded (“escrowed”) by the annuity. The risk-averse investor secures this floor income as soon as possible and, therefore, tends to favor early exercise of the option to annuitize.²⁵

The authors, however, also offer a discussion about the reasons for and merits of delaying annuitization: because risky assets carry the expectation of high return, a longer period of holding such assets “…offers people a chance to improve their budget constraint that evaporates after annuitization. So even risk averse individuals may decide to delay in the expectation of creating more wealth and enjoying a higher long-term income.” [P. 226] Individuals who (subjectively) anticipate a long life span may delay annuitization given the potential benefits of (1) lower future annuity costs—especially if interest rates increase, and (2) higher returns from exposure to risky assets.

Given the fact that an annuity purchase decision is irreversible and that the real option to annuitize has time value, the Kingston-Thorp analysis parallels the Milevsky-Moore-Young assertion that an investor will delay annuitization until such time that the expected return from the annuity contract exceeds that of other financial instruments exhibiting comparable risk. In contrast to the Milevsky-Moore-Young model which assumes CRRA, Kingston and Thorp assume Hyperbolic Absolute Risk Aversion [HARA] which is a function suggesting that an investor gains utility only for consumption in excess of either a static or fixed floor; or, for investors who ratchet up their spending targets, for consumption in excess of a dynamically changing floor. Their model assumes that the investor has a fixed consumption floor—i.e., a

²⁴ For decision making criteria incorporating expected utility and Roy’s ‘safety first’ rule, see Levy, Haim and Levy Moshe. 2009. “The safety first expected utility model: Experimental evidence and economic implications.” Journal of Banking & Finance, Vol 33: 1494 -1506. The article provides a helpful review of the literature on decision making criteria geared to maximize the probability of achieving a minimum acceptable economic result where any outcome falling short of this level constitutes an investment disaster.

²⁵ Risk aversion implies a diminishing marginal utility of wealth—the pleasure of gaining a dollar is less than the pain of losing a dollar. In this case, the pain of losing the opportunity to secure a dollar of periodic lifetime consumption is greater than the pleasure of realizing an extra dollar in each period.
The authors point out that the commonly used CRRA utility function is consistent only with a constant mix portfolio management approach. HARA utility functions however can accommodate a buy and hold approach as well as convex payoff approaches such as portfolio insurance and other dynamic asset allocation strategies. Any model using a CRRA utility function assumes that investors derive utility from consumption irrespective of its absolute level. However, it is plausible to assume that only consumption above a threshold level generates positive utility. When such a ‘non-zero consumption floor’ is introduced into the model, a HARA utility function is, according to the authors, required to solve for the optimal strategy. The presence of a consumption floor changes the decision making process in that the goal can now be expressed in terms of surplus optimization.

In the Kingston-Thorp model, the minimum standard of living target is fully funded (“escrowed”) by the annuity. The primary motivation for adopting this strategy lies in the fact that the annuity contract’s mortality premium makes it cheaper to escrow wealth earmarked to produce future income. The investor secures the floor income as soon as possible and, therefore, tends to favor early exercise of the option to annuitize. This approach to portfolio design is a variation on Tobin’s two-fund portfolio solution. Only surplus wealth is invested in the risky asset portfolio while the remainder is allocated to a risk-free annuity. “It follows that introducing a positive consumption floor has a similar effect to raising relative risk aversion. In addition, the agent recognizes that it is ‘cheaper’ to store escrow wealth in an annuity rather than a bond portfolio (at least where there are small enough loadings), creating another incentive to switch into complete annuitization at an earlier date.”

Portfolio monitoring is critical in that a key ratio is the level of surplus relative to the changing costs of securing an “acceptance level” income stream. The two-fund strategy—annuity plus risky asset portfolio—also harkens back to a 2002 work by R. Korn and M. Krekel [“Optimal Portfolios with Fixed Consumption or Income Streams”]. The reader may also recall Yaari’s observation that annuity contracts permit investors to pursue separate solutions to the retirement income and the ‘saving’ (bequest) problems.

A boundary control approach based on threshold levels of income or wealth shifts the portfolio control variable from asset allocation (long-term expected return) to the liability or spending variable. The

---

26 Technically, CRRA is nested in hyperbolic risk aversion functions. However, HARA functions encompass absolute risk aversion functions such as decreasing absolute risk aversion [DARA]—as wealth increases the investor is more comfortable committing dollars to the risky asset; increasing absolute risk aversion [IARA]—as wealth increases the investor pulls back on the number of dollars put at risk; and constant absolute risk aversion CARA]—as wealth increases the investor keeps the same amount of dollars at risk. Whenever the applicable risk metric defines the percentage of wealth put at risk, the risk model incorporates a relative risk aversion measure; whenever the applicable risk metric defines the level of dollar wealth put at risk, the risk model incorporates an absolute risk aversion measure. It is, however, the rare investor who remains indifferent to changes in the level of wealth when evaluating investment and spending strategies. This is a central criticism of incorporating a relative risk aversion measure into a utility-based model.
feasibility condition for successful asset management depends on matching the investor’s retirement assets to liabilities, where liabilities are expressed as the present value of future income, gifting and bequest objectives. Philip Dybvig’s 1999 argument that efficient asset management requires a simultaneous solution of asset allocation and portfolio distribution policies reenters the academic discussion by the end of 2005.

We also take note of a plethora of articles seeking to back test the empirical distribution of realized investment returns to discover bright-line rules to enhance the safety and sustainability of retirement income portfolios. The studies, in general, apply one or more distribution formulae to various asset allocations in an attempt to uncover the optimal combination of spending and portfolio allocation. Early studies generally apply autopilot, set-in-stone, spending rules to static portfolio asset allocations. Later articles seek to uncover flexible spending rules or dynamic asset allocation strategies that, if applied to the realized return sequence, would have enhanced retirement consumption and bequest goals during the historical period under evaluation. These articles are summarized and discussed in detail in “An Annotated Bibliography: Longevity Risk and Portfolio Sustainability [1965 – 2014]” referenced at the beginning of this literature review. ²⁷

**Additional Developments**

Commentaries and debates on many of the above-mentioned issues also appear in journals written for readers from within the estate and trust community. Although investment professionals may not often encounter these articles, they have paramount importance for asset management within a fiduciary context. In addition to the investment prudence standards embodied in federal ERISA statutes, many states have adopted versions of the Uniform Prudent Investor Act, the Uniform Principal and Income Act, the Uniform Trust Code, and the Uniform Prudent Management of Institutional Funds Act. These uniform acts, as adopted by and embodied in specific state statutes, set standards of prudence for those investing trust-owned assets in a fiduciary capacity. Trustees of irrevocable family trusts are often charged with the duty to balance the interest of current (income) beneficiaries and remaindermen beneficiaries. Thus, the discussion regarding how to produce safe, substantial and sustainable periodic income from a finite amount of capital while, simultaneously, preserving the value of terminal wealth has great import. Breaches of fiduciary duties may lead to substantial awards payable from a trustee’s personal funds. A rich body of literature reflecting a broad heterogeneity of legal opinions provides valuable additional perspective.

By the mid-2000s, the academic discussions regarding portfolio management elections split along several lines. One group of commentators advocates a “bottom-up” approach in which the investor annuitizes whenever he possesses sufficient wealth to lock in an acceptable threshold standard of living. Only surplus wealth is exposed to the risks and rewards of investing in the financial markets. By contrast, a second group advocates a “top-down” approach in which the investor delays annuitization either until an optimal stopping time—the definition of ‘optimal’ takes on various meanings depending on the financial risk metrics under investigation—or until it appears that the investor is approaching dangerously close to the free boundary. The top-down advocates espouse an annuity-as-safety-net approach. There exist, of course, a spectrum of middle-of-the-road positions which recommend gradual annuitization based on various combinations of age/wealth/income/bequest motivation triggers.

Irrespective of the suggestions flowing from a retirement income risk model, it is generally true that the greater the investment wealth relative to the demands placed upon it, the more extensive are the asset management choices. As wealth increases, as Don Ezra notes in his 2009 essay for the CFA Institute [“Who Should Buy a Lifetime Income Annuity? And When?”], a surplus provides the luxury of focusing primarily on wealth management. If a decrease in wealth makes a need to hedge longevity risk appear imminent, the annuitization option can then be considered. Academic modeling is rarely a perfect substitute for prudent judgment.

Modelling Additional Risk Factors and Input Variables

Many of the early academic papers, although mathematically elegant, construct risk models that incorporate simplifying assumptions—complete markets, a representative investor exhibiting constant relative risk aversion, a log-normal return distribution. Recent studies tend to relax many of these assumptions; and, the number of variables incorporated into risk models also grows larger. The 2008 paper by Wolfram J. Horneff, Raimond H. Maurer, Olivia S. Mitchell, Ivica Dus [“Following the Rules: Integrating Asset Allocation and Annuitization in Retirement Portfolios”], provides a comprehensive list of the important assumptions and variables that characterize risk models in the academic literature through 2007.

A good example of how retirement portfolio models expand to encompass additional exogenous and endogenous variables is the 2006 essay entitled “Life-Cycle Asset Allocation with Annuity Markets: Is longevity Insurance a Good Deal?” The authors, Wolfram Horneff, Raimond Maurer and Michael Stamos, consider: (1) the impact of economic shocks on the demand to hold non-liquid annuity contracts, (2) human capital as a non-tradable asset that is a close substitute for bonds, and (3) other economic factors. Furthermore, their model is a continuous time barrier control type problem which evaluates the option to annuitize throughout the entire life cycle as opposed to either a one-time option to annuitize at retirement or to a continuously exercisable option only during retirement.

The consequence of incorporating a broader set of investor elections into risk models is apparent in the 2007 essay by Moshe Milevsky and Virginia Young [“Annuitization and Asset Allocation”]. The authors point out that when an investor faces an all-or-nothing option to annuitize, the optimal time for annuitization of total wealth is when the expected return from the risky asset portfolio equals the risk free rate plus the annuity mortality credits. Under a CRRA assumption, “…the optimal time to annuitize one’s wealth is independent of wealth and is, therefore, deterministic.”[P. 3148] The decision rule states: when the value of the option to annuitize equals the expected value of the payoff from the underlying portfolio, then exercise the option. This is an option valuation problem. Beyond this time, any delay runs the risk of producing less future consumption than if one exercises the annuitization option immediately. The decision rule changes, however, in a model that gives an investor the capability of annuitizing a fraction of wealth at various time intervals, “…the individual’s optimal annuity purchasing is given by a barrier policy in that she will annuitize just enough of her wealth to stay on one side of the barrier in wealth-annuity space.”[P. 3139] If wealth is spent to purchase an annuity, periodic income increases. In
this risk model, the barrier’s location is where the marginal utility of annuity income equals the marginal cost of spending down wealth to secure it. When the benefit of the income exceeds the benefit of retaining wealth, the investor will exercise the option to annuitize a fraction of wealth to restore equilibrium.

Both the implied and explicit financial planning recommendations appearing in the literature are model dependent; and, with the rapid expansion of modeling capabilities, there is a marked increase in the heterogeneity of such recommendations. This puts a burden on the consumer of academic literature. The models, although quantitative in nature, do not generate conclusions having the force and effect of mathematical theorems. Rather, increased complexity puts a premium on the reader’s ability to understand a risk models’ structure, assumptions, and input variables so that he can intelligently evaluate conclusory observations. Observations which are not empirical in nature are often replete with model risk; and, a cautious investor rarely accepts them at full face value.

The 2008 study by Cassio M. Turra and Olivia S. Mitchell [“The Impact of Health Status and Out-of-Pocket Medical Expenditures on Annuity Valuation”] evaluates the utility of annuitization in the face of liquidity shocks. Even if the annuitant is in good health, “our stylized life cycle model with uncertain out-of-pocket medical expenses shows that annuities become less attractive to people facing such medical expenses.” [P. 246] This observation raises the issue of the attractiveness of annuities in an incomplete market setting: “…when both adverse selection and uncertain medical expenses are accounted for and annuity markets are incomplete, we show that annuity equivalent wealth values are fairly low for people in poor health and about 25 percent higher for people in good health.”28 [P. 246]

The economic consequences of liquidity demands generated by a deterioration in health is also the focus of the 2008 essay by Jon Ameriks, Andres Caplin, Steven Laufer, and Stijn Van Nieuwerburgh [“Annuity Valuation, Long-Term Care, and Bequest Motives”]. In contrast to the annuity-as-safety-net school of thought, the authors conclude that the demand to annuitize decreases as wealth diminishes. That is to say, there is a positive correlation between wealth and annuity demand. As financial resources shrink, the ability to fund lifestyle expenses diminishes. If an annuity purchase exhausts most liquid financial wealth, the investor incurs increased vulnerability to health shocks. At low wealth levels, a serious medical shock would simultaneously deplete remaining assets, raise current expenses, and decrease the mortality-adjusted future value of existing annuity payments. If, as the authors suggest, “retirement security…can be summed up simply as ‘having the resources you need, when you need them,’” [P. 271] then standard annuities may be only a partial solution to security in the face of severe health shocks:

28 Part Three provides a more in-depth discussion of ‘annuity equivalent wealth’
“such products do little to deal with retirees’ need for resources when emergencies arise, and they can even exacerbate financial distress in exigent situations.” [P. 272]

A risk model that suggests a decrease in the demand to annuitize as wealth decreases also finds support in the 2007 essay by Milevsky and Young, cited above [“Annuitization and Asset Allocation”]. Milevsky and Young assert that the demand for annuities is increasing with wealth, risk aversion, positive health assessment, and portfolio volatility. Gaobo Pang and Mark Warshawsky [2010] “Optimizing the Equity-Bond-Annuity Portfolio in Retirement: The Impact of Uncertain Health Expenses” argue that “…the uncertainty in healthcare spending enhances the welfare gain of life annuities, more so for higher-income households because such annuities are more likely to hedge their longer life and higher health spending.” [P. 207]

The literature review noted earlier that a large number articles focus on the question of how much can be safely withdrawn each year from a retirement income portfolio. Some studies investigating ‘safe’ withdrawal rates may mislead investors because they fail to account for a number of critical balance sheet elements. On the plus side of the ledger, investors may own life insurance for the benefit of a surviving spouse, or may expect an inheritance. The 2011 paper [“Asset allocation, human capital, and the demand to hold life insurance in retirement”] by Patrick J. Collins and Huy D, Lam provides further insight on this topic. On the minus side of the ledger are unreimbursed health costs and long-term care expenses. For example, a 2013 study [“Strategies for Mitigating the Risk of Outliving Retirement Wealth”] by Vickie Bajtelsmit, LeAndra Foster, and Anna Rappaport, estimates that health and long-term care expenses add an extra $260,000 to overall costs of retirement for the average retired married couple. Retirement is expensive.

The Bajtelsmit, Foster and Rappaport study considers households (husband age 66, wife age 63) at annual income levels of $60,000, $105,000 and $150,000. It evaluates three strategies for improving retirement income sustainability:

1. Delaying retirement
2. Cutting targeted spending
3. Acquiring Long-Term Care Insurance.

The most effective tool is, by far, delaying the start of retirement. For all households, cutting spending by 15% has only a slight impact on the risk of running out of funds. The spending reduction benefit is easily overwhelmed either by poor investment performance, or by high health costs. Although their retirement

29 If insurance proceeds are available to the younger spouse, incorporating this contingent asset into the retirement income risk model may decrease shortfall probability significantly.
income model suggests that Long Term Care [LTC] insurance produces limited benefits for the lowest income households, it is generally an ineffective tool to enhance a household’s financial security: “based solely on the probability of having wealth remaining at death and the years without wealth, it appears that LTC insurance does more harm than good.” [P. 322] Generally, “… the LTC insurance premiums increased post-retirement expenses, resulting in quicker depletion of retirement wealth.” [P. 322] The authors conclude: “… higher income families have enough wealth such that LTC costs do not play a substantial role in determining adequacy.” [P. 327]

By contrast, a paper presented by David Blanchett at the Society of Actuaries 2014 Living to 100 Symposium [“Estimating the True Cost of Retirement”] indicates that retirement cost may be significantly less than many retirement income models suggest. Blanchett looks at data from several consumer spending surveys that track expenditures for households with retired investors. The data reveal how household expenditures change, over time, at various ages. The rate of cost-of-living increases for seniors differs only modestly from the general population cost-of-living index: “… from December 1982 to December 2012, the average annual change in the CPI-E [Consumer Price Index for the Elderly] has been 3.07 percent versus 2.92 percent for CPI-U [Consumer Price Index for Urban Consumers] …” [P. 11]

Generally, according to the surveys, retired investors spend neither a constant periodic dollar amount nor a constant amount adjusted for inflation. Rather, the cost curve is convex (shaped like a smile), with expenditures gradually decreasing with age until they move upwards again towards the end of life. For example, the author cites one consumer spending study that finds “… consumption-expenditures decrease by about 2.5 percent when individuals retire; expenditures continue to decline at about a rate of 1 percent per year after that.” [P. 6] Estimating expenditure curves for households at various income levels increases the accuracy of retirement cost projections. Compared to the outputs from traditional retirement income risk models, the true cost of retirement may be less. Blanchett asserts: “… modeling the cost over the expected lifetime of the household, along with incorporating the actual spending curve, results in a required account balance at retirement that can be 25 percent less than the amount required using traditional models.” [P. 23]

More and more investors are using computer-driven, on-line programs to calculate how much they can safely spend during retirement. Likewise, investment advisors frequently employ retirement income

---

modeling software packages. If spending puts too much pressure on a retirement nest egg, the portfolio suffers a risk of depletion during the investor’s lifetime. However, as stated, when using retirement income modelling applications, the investor must also consider model risk. Model risk is difficult to detect, and, once spotted, is difficult to quantify. The Society of Actuaries and The Actuarial Foundation’s 2009 Survey entitled “Retirement Planning Software and Post-Retirement Risks” prepared by John A. Turner and Hazel A. Witte concludes: “Notably, most programs do not do a good job of evaluating the risks that users face….generally most risks are ignored.” [P. 12]

A bird’s-eye look at the literature reveals a plethora of different research methods, modeling assumptions and portfolio allocation/spending preferencing criteria, all of which may produce significantly different outputs even given the same empirical data. Conclusions are subject to model risk; and, from time-to-time, practitioners may translate the output from academic model building—an exercise designed to explore quantitative relationships among variables of interest—into prescriptive statements for investors. But the mathematical assumptions required for tractable model building often diverge in both their character and form from common investor utility functions, from the process underlying the distribution of empirical asset price evolutions, or from any realistic pattern of retiree spending decisions. Conversely, from time-to-time, practitioner-oriented articles may resort to pure empiricism in an attempt to parse historical return evolutions to find rules for safe and sustainable portfolio withdrawals.

Although historical back testing continues to appear in some recent articles, there is a growing awareness of its limitations. Articles are now more likely to employ either:

1. Shortfall probability analysis—generally derived from a Monte Carlo simulation model; or,
2. Utility-based analysis which optimizes one or more objective functions within a Life-Cycle model.

Unfortunately, the substantive differences in models outputs can confuse investors and financial advisors alike. For example, under one retirement-income modeling approach, an annuity may appear as an expensive and undesirable retirement planning instrument; another approach may characterize annuitization as a worthwhile asset management option. A heightened awareness of model risk is useful in that it diminishes the propensity to base financial strategy on any one risk measure or model type.

An awareness of risk model limitations informs an interesting essay published in 2013 by David M. Blanchett [“Examining the Benefits of Immediate Fixed Annuities in Today’s Low-Rate Climate”].

31 This study is a follow up to the 2008 Society of Actuaries monograph entitled “Managing Post-Retirement Risks: A Guide to Retirement Planning” which provides risk management tables, suggestions and comments for various types of risks faced by retirees.
Although the underlying risk model is simplistic, Blanchett’s work is noteworthy because it incorporates multiple approaches to quantifying the risks and rewards of retirement income strategies. Blanchett begins by reminding readers that the buyer of an annuity should expect to bear a cost because the annuity is a form of insurance contract. Today’s low interest rates increase the cost of periodic income because the insurance carriers cannot earn substantial amounts on their underlying bond investments. However, even in the absence of a positive expected present value, an annuity buyer may experience positive welfare.

The article presents a table of sustainable withdrawal rates derived from a Monte Carlo simulation model assuming fixed 3% inflation, constant dollar withdrawal rates based on percentages ranging from 3% to 10% of initial value, and a $1 million initial portfolio value. Planning horizons range from 20 through 40 years. For example, according to the model, at a 4% real withdrawal rate there is a 9% chance of portfolio depletion by the end of 30 years. Adding the condition of survivorship of at least one spouse over the planning horizon, drops the failure rate to 8% for a 30 year period assuming an age 65 couple.

The challenge faced by the retiring household is to determine whether the potential benefits of an annuity outweigh the costs. The author calculates annuity IRRs for a male, female, and joint couple (100% survivorship benefit) assuming they are both age 65 at the time of annuitization. The calculations indicate that IRRs for all groups become positive by approximately age 80. An early death results in a highly negative IRR, a long life may result in a modestly positive IRR. The IRR distribution has a pronounced left side skew: “From a practical perspective, the negative skew associated with IRRs…should be viewed as the ‘cost’ of offsetting the potential positive skew associated with life expectancy. The annuitant is effectively trading the possibility of dying early (and the corresponding negative IRRs) with the hedge of living a long life and having guaranteed income the entire period.” [P. 47]

The study continues as the author supplements the IRR analysis with a utility-based analysis. The model’s utility function reflects CRRA and is the standard von Neumann Morgenstern power utility function. The model defines utility maximization as a function of the total income goal replaced during retirement. This is calculated by dividing the net present value of all payments received over the retiree’s lifetime plus the total balance of assets at death, by the net present value of the total income need. This calculation, of course, considers aggregate utility across the planning horizon rather than period-by-period utility achieved by income in excess of the threshold target. However, Blanchett incorporates a term to reflect the investor’s aversion to income variance: “…the utility-maximizing portfolio will be the combination of assets that both maximizes retirement income and minimizes the downside variability associated with generating the income.” [P. 48]
The portfolio under consideration has a pre-parameterized mean of 7% with a standard deviation of 9%. This is characteristic, according to Blanchett, of a portfolio allocated 40% to stocks and 60% to bonds. Returns are lognormally distributed. Each retiree (male, female, or couple) has an inflation-adjusted pension of $30,000 per year (100% survivorship basis). If the annuity provides more than the targeted withdrawal rate, the excess is reinvested into the stock/bond portfolio. Results are tested at the currently available annuity rates and for alternative annuity pricing factors assuming interest rates 50 and 100 basis points higher. The utility-based analysis concludes: “An IFA [Immediate Fixed Annuity] is not featured for any of the current rate scenarios under age 70, and only with material allocations for those annuitizing at age 80. However, the optimal allocation to an IFA does increase at older ages, and increases considerably should IFA rates improve.” [P. 49] For example, at a 4% of initial portfolio withdrawal rate, a male should forgo annuitization until approximately age 75. At that time, 30% of the portfolio should be annuitized. “Given today’s annuitization rates, which are currently near all-time lows, many retirees are likely better off waiting until interest rates and subsequent annuitization rates improve, or delaying the IFA purchase decision to an older age. Even with today’s low rates, IFAs remain an attractive longevity hedge for retirees age 80 or older…” [P. 49] Although Blanchett’s model is rudimentary, it is instructive to note the sensitivity of normative statements not only to changes in model input values and variables, but also to modeling approaches (cost-of-retirement efficiency criteria v. utility of consumption).

Although the literature review does not cover the topic of compensation for incurring liquidity risk when contracting for an annuity, we acknowledge that the cost of illiquidity will vary from investor to investor and may be substantial. We refer the interested reader to the early studies of Amihud and Mendleson which attempt to quantify the costs of illiquidity, and to the many successor studies published in the late 1980s and thereafter.32

The subjects of model risk, modeling approaches, and portfolio shortfall probability analysis under various spending regimes receive comprehensive treatment in the 2014 Working Paper by Patrick J. Collins, Huy D. Lam and Josh Stampfli [“How Risky is Your Retirement Income Risk Model?”].33 The paper takes note of the proclivity of many researchers to present conclusions from models that pile up a series of indefensible assumptions such as a constant rate of inflation, time-invariant volatility, set-in-stone correlation matrices, Gaussian distribution of future investment returns, and so forth. On the plus side, many such papers pose interesting and important questions; on the minus, many present spurious

32 See, for example, Amihud, Yakov, and Mendleson, Haim “Liquidity and Stock Returns,” Financial Analysts Journal Vol. 42, No. 3 (May/June 1986): 43 – 48. This early article considers the time over which the bid-ask spread is amortized as required compensation over the investor’s holding period.
conclusions. Readers interested in exploring this literature are also referred to the annotated bibliography referenced at the beginning of this literature review.

Life-Cycle Models, Control Variables and Portfolio Sustainability

The utility of consumption and bequest models—especially the classic Life-Cycle models—offer valuable insight into complex interactions among longevity, asset allocation, labor income, work/leisure tradeoffs, timing of annuitization options, and portfolio withdrawal strategies (investor spending). Optimal asset allocation and portfolio management decisions vary substantially across the population of investors depending on (1) the form of the investor’s utility of wealth function—e.g., relative or absolute risk aversion, and (2) the degree of investor risk aversion—i.e., the concavity of the utility function. In general, the Life-Cycle model research shreds conventional wisdom regarding one-size-fits-all rules of portfolio design and management. There appears to be neither an optimal allocation for all seasons nor an optimal withdrawal rule for all portfolios.

Asset Management Control Variables: Retirement Spending

The amount and timing of retirement spending is, not surprisingly, an important factor in estimating the likelihood that a portfolio will be able to provide adequate lifetime income. Spending—or, more precisely, spending flexibility—is a key control variable in many retirement income risk models. For example, the 2005 essay “A Sustainable Spending Rate without Simulation,” by Moshe A. Milevsky & Chris Robinson asserts that retirement is feasible when current wealth is greater than the Stochastic Present Value [SPV] of a spending plan. A retirement plan’s success or failure is affected by age, asset allocation, and the spending target. The authors test which of these three levers of ‘retirement sustainability’ is of greatest importance in the prevention of the probability of retirement ruin. In general, they find spending rates higher than 5% of initial wealth produce unacceptably high probabilities for ruin. The spending decision dominates the asset allocation decisions at a 5%+ rate: “No matter what reasonable portfolio is chosen, asset allocation will not turn a bad situation into a good one.” [P. 97] This is because return and variance move together and any attempt to increase return also increases the failure rate. The two most effective levers for controlling retirement success, according to Milevsky and Robinson, are (1) postponement of portfolio distributions to a later age, and (2) reductions in consumption targets. This paper represents a change in emphasis from Milevsky’s earlier option valuation approach. This paper focuses asset management on monitoring wealth in excess of the SPV of consumption.

A 2006 essay by Gary Smith and Donald Gould [“Measuring and Controlling Shortfall Risk in Retirement”] tests the effect of a flexible distribution policy on shortfall probability. Their model assumes a 50% elasticity of spending—Δ10% wealth relative to initial wealth generates Δ5% spending
relative to initial spending. The flexible spending policy dominates the fixed withdrawal rule given that it produces smaller shortfall risk and higher terminal wealth across all allocations. Flexibility in spending (elasticity of consumption) is a critical factor in portfolio sustainability in this model.

Follow on studies incorporating flexible spending into a retirement income risk model include the 2008 study by R. Gene Stout [“Stochastic Optimization of Retirement Portfolio Asset Allocations and Withdrawals”], the 2008 study by John J. Spitzer [“Retirement Withdrawals: An Analysis of the Benefits of Periodic ‘Midcourse’ Adjustments”], and the 2011 study by Moshe Milevsky and H. Huang [“Spending Retirement on Planet Vulcan: the Impact of Longevity Risk Aversion on Optimal Withdrawal Rates”].

John B. Mitchell [“Retirement withdrawals: Preventive reductions and risk management”] tests, in 2011, a set of distribution rules customized to the investor’s risk/reward preferences and cash flow objectives. The application of the rules is a function of dynamically changing portfolio values as well as expected remaining lifetime of the investor. Mitchell notes: “…anticipating adverse events and taking corrective action increases our chances of avoiding larger more catastrophic problems.” [P. 45] In this case, the ‘event’ to be avoided is financial ruin. Mitchell bases his distribution rules on the ratio between the present value of an annuity calculated for the remaining expected life of the retired investor and the portfolios initial value. Additionally, Mitchell’s discount rate for calculating the PV of the annuity is the historical rate earned by the risky asset portfolio—not the rate associated with an historical, current or estimated yield curve. Under Mitchell’s approach, a shortfall probability metric trumps a utility of consumption metric.

Mitchell advocates a conservative initial consumption policy. If future portfolio returns are positive, higher portfolio value may allow for increasing dollar withdrawals. This is a start-low / go-high withdrawal strategy. Complex rules, in Mitchell’s opinion, are required because of asset price volatility. There is a maximum [MAX] and minimum [MIN] allowable withdrawal rate under all economic environments. The ongoing asset management controls focus on the percentage of calculated increase or decrease that can be actually taken in any one period: “For example, if a retiree amortizes their portfolio over their expected remaining life span at historic rates of return and finds the portfolio could sustain a 10% withdrawal rate as compared to a current 6% withdrawal rate, a 40% UR would allow them to only increase the withdrawal rate to 7.6% that is (6% + [.4 X {10% -6%}]).” [P. 49] Mitchell uses a 2006 article co-authored with R. Gene Stout [“Dynamic retirement withdrawal planning”] as a base case for comparison purposes: “Stout and Mitchell (2006) report, for example, for a 4.5% Initial Withdrawal Rate

---

34 The ratio mixes a forward-looking numerator with a backward-looking denominator.
35 Note: UR = Upward Adjustment Rate
(INIT), 2.4 UT and 1.0 DT..., 0.2 UR and 1.0 DR, 40% MAX and 3% MIN. [Note: UT and DT = Up Threshold and Down Threshold, respectively]. These controls yield a 6.63% average withdrawal rate, 4.33% probability of ruin to age 100, and averaging ending portfolio 1.07 times the beginning amount based on 1926 – 2004 data. The same controls, updated to 2008, yield a 6.47% average withdrawal rate, 8.68% probability of ruin to age 100..., and average ending portfolio 0.88 times the beginning amount. Thus, merely incorporating two years more results into the underlying data set means that there is an approximate doubling of the portfolio failure rate.....What is optimal today may not be tomorrow; presumably because of heteroskedasticity.” [P. 49] Mitchell, however, acknowledges that his rules-based system does not comfortably co-exist with a utility of consumption preference-based system: “Retirees may prefer greater consumption at younger ages when they are more active...at the expense of reduced consumption if they superannuate.” [P. 53]

The start-low / go-high withdrawal strategy recommendation is a consequence of heightened attention to sequence risk (a series of returns below expected return); and, on the consequences of sequence risk on the risk-of-ruin probability. As noted, however, conservative initial portfolio withdrawal rates are vulnerable to criticism from (1) authors of research studies incorporating a utility penalty for deviations on both the low and high side of targeted benchmark income, as well as from (2) authors employing “life cycle” utility based models which, when incorporating a factor for high subjective discounting (investor impatience), assert that optimal retirement spending usually decreases with age.

Although a review of the literature reveals that many commentators advocate a start-low / go-high withdrawal strategy, by 2012 this approach comes under full-scale assault in Gordon Pye’s book on retirement income strategies [The Retirement Retrenchment Rule]. The mundane topic of how to monitor and manage a retirement income portfolio to fund future current and future spending is becoming controversial. We devote, below, a separate section to Pye’s work.

We also call attention to a 2011 paper that Mitchell co-authors with Larry R. Frank Sr. and David M. Blanchett [“An Aged-Based, Three Dimensional, Universal Distribution Model Incorporating Sequence and Longevity Risks”] as an additional example of retirement portfolio risk model tracking a shortfall risk metric; but, concurrently, evaluating the marginal utility of income as a function of attained age. In this essay, the authors’ retirement income model focuses on portfolio withdrawals as the critical control variable. At the start of retirement, the investor sets a probability of failure rate that matches his or her risk tolerance preferences. Future withdrawal rates [WR%] are managed so that the probability of failure [POF] rate remains constant over time: “The WR% are managed so that the retiree has a target exposure to the POF rates—e.g., 5% probability of failure at age ‘x’ at a y% withdrawal rate...a set withdrawal
rate, e.g., 4%, is not optimal for all retirees because not all retirees are the same age.” [P. 9] In the authors’ opinion, “De-cumulation should be viewed as a dynamic, rather than set-and-forget, exercise.” [P. 11]

**A shift in emphasis leads to a new benchmark—and to conflicting recommendations**

The papers cited in the previous section emphasize the importance of implementing prudent spending policy to mitigate the risk of portfolio depletion prior to the end of the planning horizon. The focus on spending—especially the introduction of a threshold periodic income target—rather than on terminal wealth has a number of important consequences. A threshold income requires a change in the utility function; and, we will pick up the threads of this discussion shortly within the context of dynamically changing risk aversion. Most importantly, however, the control variable is now the retired investor’s standard of living itself. Utility and shortfall risk metrics converge to a “hurdle race” problem where the solution requires both long-term portfolio sustainability and success in providing a minimum period-by-period income. The actuarial literature outlined in Part One of this literature review commands increased attention as investigation into the techniques for and costs of producing adequate income throughout the planning horizon become the topic of the day.

At least two issues emerge: (1) a benchmarking issue—to what extent does an annuity benchmark represent a reasonable way to compare and contrast retirement income strategies; and (2) a debate regarding the prudence of when and how to incorporate annuitization in the management of retirement assets. Neither issue is new. However, the incorporation of a threshold income requirement into risk modelling changes the nature of the investigation. For example, the 2008 essay by Huaxiong Huang and Moshe Milevsky [“Portfolio Choice and Mortality-Contingent Claims: The General HARA Case”] asserts that the minimum value of income needed by the family unit over the life cycle is the main driver of the demand for either life insurance coverage or for income annuities. Sustaining a target level of periodic income is the appropriate measure of financial risk.

Commentators continue to split on whether to annuitize as soon as possible lest a forthcoming bear market jeopardize the ability to secure threshold income; or, whether to delay annuitization to capture the expected equity risk premium and, potentially, a lower cost annuity contract issued at an older age. We term the first asset management strategy the ‘annuitize ASAP’ strategy; and the second, the ‘annuity-as-safety-net’ strategy. In more stark terms, the choice is between: (1) annuitize now and resolve ambiguities surrounding the sources and amount of future income; or, (2) wait to annuitize as long as a delay remains a prudent and suitable investment management election.
A good example of the ASAP strategy is the 2007 study by David F. Babble and Craig B. Merrill [“Rational Decumulation”]. The authors suggest that a utility maximizing investor will not pursue a strategy that leaves a positive probability of failing to support a threshold level of lifetime consumption. Penetrating this minimum produces, in their opinion, infinite disutility; and, given the assumption that utility is additive across all economic states, such a strategy is irrational. Their model directs the investor to allocate risk-free assets sufficient to support the minimum periodic income goal. In a multi-period context, the risk-free asset is an inflation-adjusted annuity like Social Security. If the minimum consumption target requires periodic income greater than that available through government or corporate pension benefits, the model advises the investor to immediately annuitize a portion of current wealth to fund the deficit. Excess wealth remains invested in a financial asset portfolio. Investors annuitize up to the point where the marginal utility of an extra dollar of consumption equals the marginal disutility of spending down wealth to fund an annuity income. For investors with average risk aversion parameters, annuity loads have, in the authors’ opinion, minimal impact on the allocation decision provided that the markups above the actuarially fair price are less than 30%.

Their model is interesting for a number of additional reasons. First, it incorporates a HARA utility function by virtue of a minimum required income floor; and it assumes that the investor allocates risk-free assets sufficient to support the minimum standard of living goal. Second, in a multi-period context, the risk-free asset is an inflation-adjusted annuity. The authors argue for a ‘bottom-up’ asset management approach in which the investor, with little or no delay, converts financial assets into an annuity designed to provide threshold income. Only surplus wealth is allocated to a risky asset portfolio. This is a buy-an-annuity-and-invest-the-difference strategy. If the amount of wealth allocated to the annuity is large, however, the investor may not have sufficient remaining funds to implement the optimal allocation to the risky asset portfolio. Assuming that the risky asset portfolio has a higher expected return than the annuity portfolio, the decrease in aggregate expected return (disutility) must be balanced against reduced uncertainty in future consumption (utility). The authors point out that the feedback loop plus the wealth constraint make an analytic solution impossible.

---

36 The next section explores the topic of utility in greater detail. For now, we simply note that it is important to distinguish among models assuming static risk aversion throughout the planning horizon; dynamically changing risk aversion throughout the planning horizon, a state-preference utility function within each interval throughout the horizon; and a non-separable, non-additive utility of wealth function perhaps based on a preference for maintaining one’s historical standard of living. The key point is that different assumptions regarding the appropriate and applicable utility function lead to profoundly different preferencing criteria; and, hence, to significant differences in the ranking of investment / withdrawal strategies derived from retirement income risk models.

37 Analytic solutions are efficient and relatively easy. However, they cannot be used with multiple sources of complexity or with sequences of investment/spending decisions.
In contrast to the Babbel and Merrill study, the essay by Richard K. Fullmer, also published in 2007 [“Modern Portfolio Decumulation: A New Strategy for Managing Retirement Income”], espouses a top-down approach where the option to annuitize is a last-resort safety measure. The author asserts that systematic withdrawal plans must often reduce future spending following a bear market. However, “this amounts to managing longevity risk through spending management. This approach sacrifices the investor’s standard of living in the event of poor market returns.” [P. 1] Unlike the modern portfolio theory approach to asset accumulation where the investor is concerned with terminal wealth and where standard deviation of wealth is an appropriate risk metric, this is not the case for retirement income portfolios. Rather, according to Fullmer, a more appropriate risk measure is the sustainability of income sufficient to support a threshold standard of living. Shortfall risk, relative to this threshold standard, is a more meaningful risk metric to the investor. However, a probability measure of the likelihood of achieving a sustainable threshold is not an ideal risk measure because it fails to take into account the magnitude of failure should the threshold be breached. The author asserts that the best strategy for managing retirement income risk is to annuitize when you must—but not before. This puts him squarely in the annuity-as-safety-net camp.

The key to implementing a prudent portfolio management strategy is to evaluate continuously the option to annuitize financial assets. By exercising the option only when it is necessary to ensure a threshold standard of living, the investor takes full advantage of the time value of the annuitization option. “The key for leveraging this optionality is setting the projected cost to annuitize the investor’s desired lifetime income stream as a wealth goal in the objective function. Doing so effectively transforms longevity risk into investment risk, because now it is the portfolio’s job to preserve the ability to annuitize the desired lifetime income stream….By monitoring the investor’s wealth relative to the current cost of annuitization, the decision to invest or annuitize can be continually evaluated by a financial advisor.” [P. 6] This logic leads directly to a recommendation for “…a dynamic allocation strategy.” [P. 9] The author asserts: “…when substantial cash-flow risk is present, the objective function begins to take on more of the characteristics of a cash-flow matching model.” [P. 9] The risk management approach mirrors the hurdle race problem in which the ‘provision’ must exceed the cost of securing the threshold living standard through annuitization. The author terms this an “annuitization hurdle.” [P. 10] This threshold wealth standard is a more appropriate benchmark against which to measure the risk of shortfall than complete portfolio depletion—which measures the risk of ruin: “Portfolio values below zero represent financial ruin, while values below the annuitization hurdle represent an inability to fund the desired annuity.” [P. 10]
Under this risk-management approach, the investor monitors the cost of buying an annuity to fund threshold income and compares this cost to the market value of assets remaining in the portfolio. The decision becomes how much of the portfolio surplus to put at risk before exercising the option to annuitize. Fullmer also provides a refreshing counterpoint to the conventional wisdom that suggests a retiree with dwindling resources must assume more risk in the hope that the portfolio will bounce back from bear market declines.

We note, additionally, that Fullmer’s opinion regarding the drawbacks of shortfall risk measure predates arguments made by Michael Kitces in his 2012 blog post [“Is the Retirement Plan with the Lowest ‘Risk of Failure’ Really the Best Choice?”]. Kitces asserts: “…the plan with the lowest risk of adjustment may not be the ideal plan for the client…” [P. 1] A plan with a relatively low risk of adjustment may require a draconian adjustment conditional on the adjustment requirement. Alternately, a plan with a higher risk of adjustment may require a less disruptive change in retirement income: “In other words, it may be better to follow the plan that leads to a slow failure – which can be easily fixed with mid-course adjustments – than a fast failure.” [P. 3] Kitces, like Fullmer, believes that the investor must consider both the risk of adjustment and the potential magnitude of the adjustment.

Finally, we call attention to the 2007 essay [“Integrating Optimal Annuity Planning with Consumption-Investment Selections in Retirement Planning”] by Aparna Gupta and Zhisheng Li. The article contributes to the annuity-as-safety-net approach to portfolio management. High levels of wealth diminish the demand for annuitization of financial assets. Additionally, their model locates an upper bound for the age of annuitization because of ‘brevity risk.’ This means that the length of the planning horizon becomes a factor in the decision to annuitize. A sudden change in health may make an annuitization decision have poor results—i.e., consumption could not be accelerated—which would decrease investor utility. It is interesting to note how the previously cited 2011 study by Frank and Blanchett [“An Aged-Based, Three Dimensional, Universal Distribution Model Incorporating Sequence Risk’”] further develops the Gupta and Li insights into an aged-based utility of consumption portfolio preferencing criterion.

**Utility—again**

Each investor has a unique utility of wealth function. It is often convenient to employ a power utility function with the most common expressions using either an exponent of 0.5—quadratic utility; or .0001—logarithmic utility. Thus, an investor may express the value of $100 as the square root of 100, (10 utiles) assuming a quadratic curve in wealth/utility space; or as the natural log of 100, (4.61 utiles)

38 Log utility’s exponent approaches zero at the limit. An exponent of one signifies a risk-neutral investor. Quadratic utility is similar, but not identical, to mean-variance utility.
assuming a logarithmic curve. Each exponent value generates a curve with differing slope values across the wealth spectrum. The first derivative of the slope at each point along the curve measures the marginal utility of a unit of money at that point. With the exception of gamblers’ utility curves, a utility curve is concave with a positive first derivative (more wealth is better) and a negative second derivative (the pain of a dollar loss is greater than the pleasure of a dollar gain). The slope value, therefore, measures an investor’s sensitivity to gains or losses at any given wealth level. When wealth is low, the marginal utility of an extra dollar is high; and, vice-versa. Risk aversion is the reciprocal of the marginal utility of wealth, and is usually incorporated into risk models with a negative exponent value to reflect the fact that it occupies a position in the denominator of the fraction.

An economic interpretation of the first and second derivatives is as follows: the first derivative indicates how the investor’s relative preference (marginal utility) for differing levels of wealth changes as wealth increases or decreases; the second derivative indicates the extent to which investor’s marginal utility itself is changing as wealth increases or decreases. Thus, risk aversion is the elasticity of marginal utility with respect to wealth.

Students of finance are familiar with risk aversion curves primarily through the ubiquitous illustrations of Harry Markowitz’s technique for locating the optimal portfolio at the intersection of the efficient frontier and the highest investor indifference curve. Indifference curves are risk aversion curves. The steepness of the indifference curve determines its point of tangency with the set of efficient portfolios; and, the steepness, in turn, is a function of each investor’s risk aversion. In the case of mean-variance utility, the optimum is achieved by finding the highest feasible indifference curve. When either utility of wealth (the ability to make gifts and bequests) or the utility of consumption (the ability to meet periodic income targets throughout the planning horizon) is an important preferencing criterion, it is of utmost importance to recognize that model outputs may be extremely sensitive to the nature and form of the implied utility / risk aversion function(s).³⁹

It is useful to distinguish among models that assume:

- static risk aversion throughout the planning horizon;
- dynamically changing risk aversion throughout the planning horizon;
- a state-preference utility function within each interval throughout the horizon; and,

a non-separable, non-additive utility of wealth function perhaps based on a preference for maintaining one’s historical standard of living.

The key point is that different assumptions regarding the appropriate and applicable utility function lead to profoundly different rankings of investment / withdrawal strategies derived from retirement income risk model outputs.

Many utility-based retirement income models assume a Constant Relative Risk Aversion [CRRA] function for several reasons:

1. Such a function assumes that the investor is willing to invest a constant fraction of wealth in a risky-asset portfolio irrespective of the actual level of wealth at any moment in time—i.e., CRRA is ‘wealth invariant.’ This greatly simplifies the mathematics required to achieve a solution to the risk model.

2. CRRA approximates, under a range of conditions, mean-variance utility under which an investor cares only about the first two moments of the distribution of returns. This allows the model builder to assume a normal distribution of returns, which, in turn, also simplifies the mathematics.

3. When only a return distribution’s first two moments are portfolio choice factors, the Sharpe Ratio becomes a permissible performance evaluation metric and the optimal weight in the risky asset is given by the Merton Optimum.

4. CRRA utility allows for the application of classic von Neumann-Morgenstern-Savage axioms of utility. For any risk aversion value, the model builder can compute the wealth certainty equivalent. This permits a cardinal ranking of asset management approaches from models generating aggregate consumption and terminal wealth utility values.

Unfortunately, although the CRRA function has many useful modeling properties, there are a variety of well-known shortcomings. As noted, the introduction of a floor or threshold amount of wealth or income demands functions like the Hyperbolic Absolute Risk Aversion [HARA] or a State Preference Utility function. State preference utility assumes an individual wishing to avoid low consumption opportunities during bad economic periods. A dollar in a recessionary economy is more highly valued than a dollar during a period of prosperity. Similarly, when a retiree is interested in smoothing consumption over time, the model builder might wish to distinguish risk aversion in a single period from intertemporal risk aversion by using an Epstein-Zin utility function. The Epstein-Zin function uses an additional term to represent the elasticity of intertemporal substitution [EIS]. Highly risk averse investors have a low value of EIS because they wish to avoid large swings in consumption from one period to the next. A critical risk metric to track in models using Epstein-Zin utility is multi-period consumption variance.
Behavioral theories of finance posit that investors exhibit loss aversion. This can introduce convexity into domain of the utility function as wealth drops beneath a critical value or reference point. In addition to the family of utility functions described by Venter in 1983 [“Utility With Decreasing Risk Aversion,”], a variety of other utility functions including ‘Fisher utility’ and ‘habit’ or ‘standard of living utility’ are also found in retirement income risk models.

A good example of a study that compares conclusions based on a model assuming CRRA utility with conclusions derived from other utility functions including habit utility is the 2005 study co-authored by Nobel Prize winner Peter Diamond with Thomas Davidoff and Jeffrey Brown [“Annuities and Individual Welfare”]. They begin by pointing out that the classic Yaari Life-Cycle consumer with no bequest objective and with an uncertain date of death annuitizes all wealth under the assumption that the consumer is an expected utility maximizer with intertemporally separable utility, and with access to actuarially fair annuity products. The authors, however, contend that it is not necessary to assume that the consumer is an exponential discounter, or that he obeys the standard von Neumann-Morgenstern-Savage utility axioms, or for the annuity to be actuarially fair. Consumers, under a Yaari-like model, will annuitize all wealth provided that they have no bequest motives and that the net rate of return on the annuities is greater than the return on conventional assets of matching financial risk.

Under models incorporating other utility functions, however, immediate annuitization may not be optimal. The authors present a simulation model for a single male, age 65. Their model assumes a power utility function, exponential discounting at a deterministic rate, and general population mortality. They calculate four welfare measures after applying the model to investors with both separable utility and standard of living utility:

- Increase in wealth required to hold utility constant while moving from a constant real annuity to conventional bonds;
- Fraction of wealth optimally committed to real annuities instead of bonds;
- Increase in wealth required to hold utility constant while moving from optimal annuity position to conventional bonds; and,
- Gain in utility from selecting the optimal payout trajectory (bonds or annuities) with no requirement for either.

The assumption regarding market completeness turns out to be especially important in the decision to purchase an actuarially fair annuity. Annuities are illiquid and may not generate sufficient funds in the

---

40 Utility of consumption in period two does not depend on the investor’s consumption (standard of living) in period one. Each period’s utility can be estimated separately with aggregate utility the sum of all separately evaluated periods.

41 An investor with intertemporally separable preferences, log utility, and a discount rate equal to the risk-free rate.
event of a liquidity crisis that may arise from emergencies like uninsured medical care expenses. Consumers are willing to commit to a fixed plan of expenditures at a starting time only if they are able to trade goods across all periods and all states of nature (i.e. trading is a means of reversing or revising initial decisions). But an annuity, by its very definition, is an irreversible contract. Thus, the standard way to assess an annuity’s benefit overstates its utility value. Generally, the authors’ model suggests that the greater the wealth, the lower the demand to annuitize. The incompleteness of markets may render annuitization of a large fraction of wealth suboptimal. In terms of the ASAP v. safety-net debate, the Davidoff, Brown and Diamond model provides support for delaying annuitization, if prudent.

Furthermore, according to the authors, some studies do not capture the utility loss from locking in a fixed periodic income. In one sense, annuities are costly because they contribute to consumption constraints: “The welfare effects of larger increases in annuitization are more difficult to sign because they may constrain consumption.” [P. 21] Full annuitization may distort consumption (place an upper bound on feasible future consumption) and, therefore, may not be optimal. This is especially the case whenever utility is measured relative to the level of past consumption (standard of living). The utility of living in a studio apartment is different for a person who has lived in one throughout their life than for someone who lived in a mansion during previous years. The article contends that investors trade off consumption between periods based not only on budget constraints, but also on standard of living ratios. Marginal utility of consumption in any period incorporates two effects not present in the additively separable utility model:

1. The effect of present standard of living on present marginal utility; and,

2. The effect of present consumption on future period utility through subsequent standards of living.

The authors conclude that the interaction of various model components—the form of the utility function, the assumption of complete markets, and the extent of budget constraints—make it difficult to estimate the welfare benefits of annuitization.

The 2008 essay by Huaxiong Huang and Moshe Milevsky [“Portfolio Choice and Mortality-Contingent Claims: The General HARA Case”] also abandons a strict reliance on a CRRA function in favor of modeling preferences using a HARA function. The HARA function is appropriate for a model generating utility values only at or above the floor consumption targets. Indeed, the authors concede that maximizing utility over all investment outcomes is less than ideal; and they note that using a state-dependent utility calculation would be more realistic. This observation mirrors, in some respects, the issues raised in “The Hurdle Race” essay cited earlier.
Kim Balls, in 2006, provides a model featuring a changing risk aversion parameter as a function of changing health states [“Immediate Annuity Pricing in the Presence of Unobserved Heterogeneity”]. Ball’s utility-based model assumes discounting of contingent future consumption—i.e., the model weights the utility of consumption by the probability that the investor is alive to enjoy it. She develops an annuitant mortality model: “optimal consumption decision is derived for the case where no annuity market exists. Using the same utility framework, we derive the threshold price for an immediate annuity.” [P. 104] The model, however, focuses primarily on the health state of the potential annuitant as opposed to chronological age. The health state in any period can improve or deteriorate according to a transition probability modeled by a Markov transition matrix. Transition probabilities are calibrated by maximizing a log likelihood function from the data in the U.S. Census population mortality tables for the years 1900 – 1990.

The paper incorporates the health transition process into an economic model in which the utility of consumption in each period is influenced by the annuitant’s health state. Bequests are not considered. Per-period utility, conditional on health state j is defined as:

\[
U_j(c_t) = K_c + (\alpha + 1)^{-1}(\theta_jc_t)^{\alpha+1}
\]

Where c is consumption, \( \theta \) is a health-state consumption modifier, \( \alpha \leq 0 \) is the risk-aversion parameter, and K is a utility constant whose sole purpose is to make utility positive. The author acknowledges that the choice of the utility function (constant relative risk aversion) “…is based on mathematical simplicity…Relative risk aversion (risk aversion divided by consumption or wealth) is constant.” [P. 108]

In a market where the consumer has access to annuities, “the value of wealth is simply the utility value of the annuity purchased with the assets.” [P. 111] The model solves for the payout rate that the annuity must provide so that the utility of annuitized wealth exactly equals the expected utility of lifetime consumption given the investor’s current health state, risk aversion coefficient, and time preference discount rate. For example, an investor in the best health state, with a risk aversion coefficient of -1.50 and a 2% time discount preferencing rate requires an annuity payout rate of 5.41% per year when the expected rate of return on invested assets is 5%. A person in the worst living health state is indifferent to maintaining the portfolio with an expected return of 5% or annuitizing wealth with a payout rate of 11.88%. Changing the risk aversion coefficient to reflect greater risk aversion changes the indifference rates: “…the higher the risk-aversion parameter, the greater the policyholder’s premium for retirement risk protection.” [P. 111] Changing the health-state consumption modifier to a state-dependent variable indicates that a constant annuity payout is no longer optimal. The optimal annuity has extra payments as lower health states emerge. The model confirms that there “…are advantages to delaying annuitization, particularly when market returns available to the policyholder are superior to those available in the form
of an annuity. However, the effect here is heterogeneous, depending also on the expected longevity of the policyholder.” [P. 113]

These findings mirror those of Davidoff, Brown and Diamond (cited above) who find that the incompleteness of markets may render annuitization of a large fraction of wealth suboptimal. The Balls article is also significant in that it expands the definition of risk aversion to include a factor for what Gupta and Li, in 2007, call ‘brevity risk’ [“Integrating Optimal Annuity Planning with Consumption-Investment Selections in Retirement Planning”]. The literature survey makes apparent several points:

1. Realistic modeling of the dynamics of retirement income portfolios is becoming increasingly complex;
2. Prescriptive advice changes significantly with the choice of input variables, with the form of the assumed utility function, and with assumptions regarding stochastic modeling of investment return and inflation processes.42

New Definitions of Prudent Asset Management

Modeling retirement income portfolios under the stress of fees and distributions rapidly moves towards a consideration of asset-liability management. Incorporating a wealth-invariant CRRA function operating over the entire wealth domain becomes increasingly problematic. Many commentators begin to define prudent asset management as a function of the current wealth level rather than a function of an optimal time to exercise annuitization options, asset management rules-of-thumb (e.g., loading for equity improves long-term portfolio sustainability), or bright-line rules for retirement spending (e.g., the 4% of initial portfolio value rule adjusted for subsequent changes in inflation).

A good example of the shift in approaches to retirement income modeling is the 2006 essay by Russell Gerrard, Steven Haberman and Elena Vigna [“The Management of Decumulation Risks in a Defined Contribution Pension Plan”]. This paper updates their 2004 research cited above. They assume that current wealth is insufficient to purchase an annuity at the desired level of consumption and that the retiree elects to invest in risky assets with the hope of achieving a more favorable future income stream. There is a subtle, yet important, rephrasing of the investment issues. The problem is now expressed as a risk-ruin-to-achieve-future-wealth-goal where the objective is to maximize the probability of attaining the goal while minimizing the probability of bankruptcy. The question, of course, is whether this is a prudent strategy. The authors argue for constant monitoring of fund size relative to its ability to support

performance targets. Their normative model produces recommendations mirroring the ‘hurdle race’ asset management approach discussed above. In the authors’ opinion, risky asset positions should be maintained during times of a shortfall in wealth. Over time, the shortfall is “cured” by continued exposure to risk. This recommendation contrasts with that of Milevsky & Robinson [“A Sustainable Spending Rate without Simulation”] which shifts the control variable for portfolio sustainability from asset allocation to the timing and magnitude of portfolio distributions. We also take note of Sid Browne’s conclusion [“The Risk and Rewards of Minimizing Shortfall Probability,” 1999] that the risk of a risky asset position must increase as the time available to correct a shortfall decreases.

The reader is also referred to a 2012 paper by Russell Gerrard, Bjarne Hojgaard and Elena Vigna [“Choosing the Optimal Annuitzation Time post Retirement”] which uses the term “annuity risk” to characterize the distribution of the Present Value of an annuity. This is the risk that lower future interest rates may increase the cost of the annuity to the point where the periodic income is less than the currently achievable annuity income. Wealth falls into a region of “continuation” in which the investor does not annuitize until entering into a “stopping region” when risky assets are then converted into annuity income. The distinction is between an optimal (“propitious”) time to exercise the option to annuitize which is based on ‘financial convenience’ and a necessary time to annuitize which is based on the level of portfolio wealth.

**From Fixed Asset Allocations and Bright-line Spending Rules to Dynamic Risk Monitoring**

Sustainability of adequate lifetime income is a critical portfolio objective for retired investors. Commentators often define sustainability in terms of (1) a portfolio’s ability to continue to make distributions throughout the applicable planning horizon, or (2) a portfolio’s ability to fund a minimum level of target income at every interval during the planning horizon. The first approach focuses on the likelihood of ending with positive wealth, or, if wealth is depleted prior to the end of the planning horizon, on the magnitude and duration of the shortfall; the second focuses on the likelihood of consistently meeting all period-by-period minimum cash flow requirements.

Depending on the structure of a retirement income risk model, conclusions about cash flow sustainability are usually reached by determining the likelihood that distributions (fixed amounts, percentage of corpus, or “dynamic”) can be maintained for either deterministic or stochastic time periods under various asset allocations and longevity assumptions. ‘Sustainability,’ which is a forward-looking risk metric, is usually quantified by a projection produced by simulation-based models. It differs from the concept of ‘feasibility.’ Feasibility is a ‘solvency’ calculation. One way to ascertain financial solvency is an actuarial calculation to determine if a retirement income portfolio is technically solvent—i.e., the current
market value of assets equals or exceeds the stochastic present value of the cash-flow liabilities. If the value of assets is less than the cost of a lifetime annuity, the targeted periodic distributions exceed the resources available to fund them. Under the annuity cost benchmark, such a portfolio violates the feasibility condition. We note that determination of the feasibility of retirement income objectives is not subject to model risk because the determination rests on current observables—annuity cost vs. asset value—rather than projections of financial asset evolutions and investor longevity.

Portfolio insolvency is not the same as portfolio depletion. A portfolio is depleted when it runs out of money. However, even a portfolio holding assets with a large current market value is technically insolvent if the asset value is less than the cost of funding future liabilities. Avoiding insolvency may require a different approach to portfolio monitoring and management than one focused exclusively on either maximization of utility or minimization of shortfall probability. The trend in recent academic literature is away from building models that assume CRRA utility, normal distribution of asset returns, time-invariant volatility and correlation parameters, constant inflation, and set-in-stone withdrawal formulae. Indeed, it is somewhat surprising that the practitioner-oriented literature continues to produce a multitude of articles seeking optimal spending and asset management strategies derived from portfolio models embracing the above-listed assumptions.

The focus on the cost of retirement elevates the importance of monitoring financial resources both in terms of changes in investor needs and circumstances and in terms of the portfolio’s financial ability to discharge evolving liabilities. How does the investor assess the portfolio’s current financial condition? What is a suitable liability benchmark? What is the likelihood that, absent changes in spending and asset allocation, the investor may have to make future changes in his or her standard of living? Conditional on a shortfall in future resources, what is the likely range of shortfall magnitude and length of its duration—i.e., the distribution of time-alive-but-broke? These and other interesting questions constitute the subject matter of recent research into the areas of longevity risk and portfolio sustainability.

The need for effective portfolio monitoring as a precondition for prudent asset management finds support in the previously cited essay by Milevsky and Huang [“Spending Retirement On Planet Vulcan: The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates”]. The article contains a valuable discussion concerning the “rational reaction to a financial shock” [P. 51] over the retirement planning horizon. A spend-4%-of-wealth rule provides no guidance on updating spending policy in response to a shock to wealth resulting from a severe market decline. The authors recommend the following steps:

1. Recalculate the retirement income risk model from time zero, but with current wealth equal to the lower amount, and compute a new wealth depletion time equation.
2. Compute the new optimal consumption equation which yields an amount that differs from the pre-shock amount.

3. Continue retirement consumption at the new amount.

The study provides an example of an age 70 retiree experiencing a 31% decrease in portfolio value that necessitates an approximately 20% decrease in consumption. The rational reaction to shocks is “non-linear and dependent on when the shock is experienced as well as the amount of pre-existing income…” [P. 54]

The need for reassessment of investment strategies as a result of market declines also finds expression in a number of essays appearing in 2010. These include (1) “Freedom at 55 or drudgery till 70?” by Nabil Tahani and Chris Robinson: “Telling a client the standard deviation of returns utterly fails to portray the risk of falling short of a goal.” [P. 276] (2) “Issues in the Issuance of Enhanced Annuities,” by Robert L. Brown and Patricia L. Scahill: “…wealth relative to living expenses is an important factor in the individual’s ability to self-insure the longevity risk.” [P. 9] (3) “Spending Rates, Asset Allocation, and Probability of Failure,” by James L. Davis. Davis cautions against trying to develop a set of autopilot portfolio distribution rules: “Rather than trying to develop such a mechanical rule for time-varying parameters, it may make more sense for investors (and their advisors) to periodically review the appropriateness of current parameter values. This would allow the individual’s current circumstances to be factored into the decision.” [P. 19]

**What Does Retirement Cost? An Annuity Benchmark**

Many retirement income models peg the cost of retirement as the stochastic present value [SPV] of the liabilities—income, gift and bequest—that the portfolio must fund. Indeed, the feasibility condition, which motivates ongoing tracking of solvency status, requires that the current market value of assets equals or exceeds the liability’s SPV.

Additionally, one commonly encounters the following portfolio evaluation approaches within the literature:

- A back test of historical returns to determine suitable combinations of spending rules and asset allocation policies.
- A bootstrapped reshuffling of historical returns.
- A Monte Carlo simulation of predetermined parameter values to evaluate how various allocations can withstand the stress of portfolio withdrawals.
Horizon estimates can be simplistic—e.g., life expectancy; pre-set—e.g., 30 years; or, can reflect either the force of mortality operating over the general population—e.g., Social Security mortality tables—or operating over the population segment of high net worth investors.

Given the empirical fact that, until recently, the U.S. S&P 500 index has regularly outperformed bond indexes over periods greater than ten years, many studies recommend maintaining an equity position greater than that suggested by the financial planning rule of 100 minus current age. Prior to the 2008 – 2009 global recession, an often-heard recommendation was ‘load for equity.’ However, the precipitous drop in stock values underscored the fact that a retired investor’s standard of living depends on actual rather than expected return. Given the uncertainty in mean and variance parameters, how confident can the long-term investor be with analytical approximation formulae, empirical back tests, or simulation outputs that advocate a stocks-always-save-the-day approach—especially when retirement distributions create path dependencies (sequence risk)? Questions of (1) a credible measure of retirement’s true cost; and, (2) a credible methodology to assess that the portfolio owns sufficient assets to cover the cost, assume greater urgency post-recession.

For example, a short 2011 essay by Moshe Milevsky [“What Does Retirement Really Cost”] contends that an immediate annuity is the best measure of retirement cost. Milevsky insists that the cost of providing adequate retirement income is not magically reduced by loading a portfolio with higher expected return assets: “Enter the retirement planning software used by confused—or unscrupulous—financial advisors and they seem to offer a better and more soothing answer. If you invest more aggressively then you don’t have to use the small, pathetic and depressing 1.5% real return….” [P. 3] However, this is a mirage—“You can’t tweak expected return (a.k.a. asset allocations) assumptions until you get the numbers that you like.” [P. 4] This is a direct refutation of a ‘what’s-my-number?’ approach. Stock returns are uncertain and “pricing” the cost of retirement based on expected stock returns is the equivalent of making a bet: “…assuming a more aggressive rate of return…and then claiming that retirement has suddenly become ‘cheaper’ is a dangerous fallacy that will end up costing many retirees quite dearly.” Milevsky asserts: “…the annuity price is actually a market signal of what retirement really costs.” [P. 5]

Although financial economists had long recognized the importance of portfolio monitoring and surveillance policy for prudent asset management, researchers began to pose questions about the risk metrics on which such a policy should focus. What are the best metrics for assessing portfolio risk and what are the asset management elections available to investors concerned with a financially successful retirement? The study co-authored by Nabil Tahani and Chris Robinson [“Sustainable Retirement
Income for the Socialite, the Gardener and the Uninsured” combines risk metrics—shortfall risk + feasibility—to generate insights into retirement consumption elections. The essay notes: “retirees can and often do adjust their spending to some extent to respond to changes in their endowment due to higher or lower than expected investment returns.” [P. 188] Tahani and Robinson argue that spending adjustments are the most important factor in reducing the risk of shortfall. They observe that, rather than a fixed consumption policy, retirement spending often follows a declining pattern. The sustainability of the portfolio depends on both the initial endowment and the spending pattern that unfolds over the retiree’s lifespan. The authors state that most research on portfolio sustainability either presents a series of *ad hoc* rules for spending change or assumes a constant amount of real consumption. By contrast, Tahani and Robinson treat consumption as a stochastic variable with a drift component of \(-\alpha\), and a volatility of \(\beta\) (a geometric Brownian motion process). When \(\alpha\) is a positive number, the \(-\alpha\) drift represents an exponential decline in consumption. Further, their retirement income risk model correlates consumption to the portfolio’s real return thus making consumption stochastic. When the stochastic present value [SPV] of consumption is greater than portfolio value [wealth], there is a positive probability of ruin. They conclude: “The most significant effect on probability of shortfall … is the different patterns of consumption.” [P. 193] As previously noted, David Blanchet’s paper presented at the Society of Actuaries Living to Age 100 conference [“Estimating the True Cost of Retirement”] estimates retirement cost given empirical spending patterns of retired investors.

In contrast to Tahani and Robinson, Feng Li argues the importance of asset allocation. His argument, however, emerges from a different context. Li is interested in exploring the future distribution of annuity costs and in comparing this distribution to the distribution of returns achievable by self-annuitizing with various portfolio asset weightings. Li’s 2008 study [“Ruin Problem in Retirement under Stochastic Return Rate and Mortality Rate and its Applications”] is of great interest in that it is one of the few that explores, in depth, the distribution of the present value of a life annuity under stochastic interest rates and mortality. Assuming that current wealth permits the purchase an annuity providing sufficient periodic consumption [i.e., current wealth \(\geq\) PV annuity], then annuitization is a risk-free strategy in terms of locking in a nominal consumption floor. Electing a self-annuitization strategy, by contrast, carries a positive probability of ruin. Thus, the PV of an annuity is an appropriate standard for measuring the risks of self-annuitization.

Written as a Master’s Thesis for the Department of Statistics and Actuarial Science at Simon Fraser University, Li’s work provides an intellectual underpinning for a credible retirement portfolio monitoring and surveillance methodology. The study makes the case that the best measure of self annuitization risk is the uncertainty surrounding a portfolio’s ability to maintain a value equal to or greater than the present
value of an annuity. The risk of continuing to manage a risky-asset portfolio in the hope that it will produce wealth sufficient to buy an annuity generating higher future income is, in part, a function of the distribution of future annuity costs.

Li defines the term ‘present value of an annuity’ as the present value of the cost of funding lifetime retirement obligations either through direct purchase of an annuity contract, or through a strategy of self-annuitization. Thus ‘annuity’ can, depending on the context in which the word appears, refer to a contract offered by an insurance company or a sum of money necessary to provide withdrawals from an investment portfolio. The annuity contract benefits the retiree because of its embedded mortality premium; the risky asset portfolio benefits the retiree because of its expected risk premium. Assuming a wealth-to-consumption ratio \([w/k]\) that must finance constant lifetime consumption \((k, \text{ where } k \text{ matches the withdrawal amount provided by a commercial annuity})\), the probability of ruin (portfolio depletion) is the likelihood that the PV of lifetime consumption exceeds the PV of the annuity alternative.

Li’s procedures for generating the interest rate evolutions required to price annuity contracts in the future years employs an Ornstein-Uhlenbeck [OU] process for generating stochastic interest rate evolutions. OU is a mean reverting process where the instantaneous change in interest rate [the force of interest] is a differential equation with terms for (1) a coefficient of reversion, (2) the magnitude of difference between current and long term average interest rates, and (3) a diffusion coefficient \([\sigma]\) applied to a standard Brownian Motion process. An OU process allows for autocorrelation in both the drift and diffusion terms. As the OU process unfolds over time, the rate of return accumulation function is the integration from time zero to time ‘\(t\)’ of the investment returns. The return accumulation function \([Y(t)]\) is normally distributed and the present value of the accumulation function is lognormally distributed. Finally, the PV of the annuity function is the standard actuarial annuity pricing formula with its terms adjusted for uncertainty in both life expectancy and in the interest rate. In continuous time finance, the expected price of a whole life annuity (the first moment) is \(\int_{0}^{\infty} E(e^{-Y(t)}) P_x \, dt\). The discrete time equivalent uses sums rather than the integral with the limiting upper bound as the oldest age in the mortality table. Li calculates the first four moments of the PV annuity function.

The key element for an effective retirement income portfolio monitoring system lies in estimation of the distribution of the present value of an annuity. Li employs three methods to calculate this distribution: (1) a recursive formula (uses trapezoidal numerical integration), (2) fitting the moments to known distributions (best fit = reciprocal gamma distribution); and (3) simulation under an OU process. The simulation and recursion results are close; the gamma distribution has a poor left-tail fit especially for high volatility parameters. The study concludes that the choice of asset allocation is the single most
important factor impacting the probability of ruin under a measure that uses the annuity payout as the benchmark for the withdrawal amount.

We also call attention to the 2009 paper authored by S. Gowri Shankar [“A new strategy to guarantee retirement income using TIPS and longevity insurance”] as an example of establishing a retirement cost benchmark based on combinations of TIPS (Treasury Inflation Protected Securities) and annuities. In this case, the author claims that alternative investment strategies must be evaluated not solely in terms of their expected future returns or their shortfall probability under specific withdrawal strategies, but rather, relative to a risk-free benchmark matching the investor’s minimum target income. After stating that an annuity is the only instrument that eliminates the possibility of ruin, the author proposes an Inflation Protected Retirement Annuity [IPRA] strategy unfolding in two stages:

1. For the initial stage, purchase of TIPS until a target age is reached (e.g., age 80). The TIPS portfolio is depleted at this time.
2. At the time of the TIPS portfolio purchase, a premium is paid to buy a deferred annuity that pays out only in the event that the investor survives stage one. For surviving investors, the annuity pays a constant dollar benefit for life. A constant dollar annuity option, however, is not currently offered in the marketplace and investors might need to adjust the future annuity payout to reflect the erosion of nominal payouts over time.

Shankar’s recommendations parallel those made by Jason S. Scott in his 2008 essay entitled “The Longevity Annuity: An Annuity for Everyone?” The importance of an annuity benchmark increases as authors begin to re-think the appropriate set of risk metrics. We cover this topic in the next section

**New Characterizations of Risk Metrics**

Initially, many retirement income models tested primarily for shortfalls in terminal wealth under various pre-determined spending policies. Commentators often advocate a high portfolio loading to equity under the assumption that high expected return assets can best mitigate longevity risk. The prescription to (1) hold stocks, (2) stay-the-course, and (3) keep consumption low flows primarily from a focus on a terminal wealth shortfall risk metric—e.g., a safe retirement portfolio should be expected to last 30 or more years given a healthy retiree age 65. The implication is that the prescription to load for equity is prudent because portfolios tilted towards fixed income present an unacceptably high risk of ruin.

Some recent modeling efforts focus on both terminal wealth—wealth remaining after a pre-set number of years, or wealth remaining at the end of the investor’s life—and on the periodic income available to support consumption targets. When the retirement risk model focuses primarily on terminal wealth, the control variable is spending. Unfortunately, this does not comport well with how many retired investors
deal with risk—especially in light of the need to support a threshold standard of living. This section provides a brief review of an expanded set of risk metrics. It offers examples of models designed to provide insight into wealth v. consumption tradeoffs central to prudent retirement portfolio monitoring and management.

The importance of providing adequate periodic income augments the importance of annuities in retirement portfolio modeling and monitoring. The enhanced role that annuities play in the discussion reflects both (1) recent recommendations for establishing actuarial guidelines—in some models the guidelines assume the status of a benchmark—for assessing the portfolio’s financial condition relative to outstanding income, gift, and bequest objectives, and (2) for considering acquisition of an annuity to provide future income through contractual rather than investment means.

**Terminal Wealth v. Consumption Variance**

Rather than focusing exclusively on the probability of a shortfall in terminal wealth, the 2011 essay by Duncan Williams and Michael Finke [“Determining Optimal Withdrawal Rates: An Economic Approach”] tests portfolio design and consumption policy under the assumption that: “…the appropriate portfolio allocation in retirement is the one that minimizes consumption variance given the chosen withdrawal rate.” [P. 36] A high withdrawal rate increases the probability that the portfolio will run out of money with the result that the client must live only on “non-portfolio income.” The authors acknowledge: “…the more that is consumed from portfolio withdrawals in retirement, the higher the variance due to increased shortfall risk.” [P. 37] However, rather than defaulting to an exclusive use of a shortfall risk metric, the authors advocate “A more holistic approach to distribution planning [which] would attempt to design a distribution strategy that optimizes consumption given the strategy’s shortfall risk and client’s aversion to [income] variance….” [P. 37] Their retirement income risk model assumes that the investor has sufficient funds outside of the portfolio—e.g., Social Security, and other pension wealth—so that portfolio depletion is not deemed to be a catastrophic financial event. Their model incorporates the CRRA form of the utility function.

In contrast to studies that emphasize a host of terminal wealth risk metrics—ratio of ending wealth to initial wealth, nominal or constant dollar remaining wealth, and so forth—the authors point out that longevity risk may result in a time-preference discounting rate that differs from the general risk-free rate: “…a utility maximizer will discount future consumption based on the probability of being alive for each year in the future. If the discount rate is 4% per year, the expected utility from consumption at age 82 will be only approximately half the utility from consumption at age 65. In order to maximize expected lifetime utility then, a retiree would consume more in the early years of retirement and less in the later years when the probability of being alive is lower.” [P. 39] Furthermore, “A person might be incented to
defer some consumption until a later time if the expected return were high enough. Both the discount and expected return rates are inversely related to risk aversion. A person with a high RRA will have a relatively low expected rate of return due to conservative portfolio choice, but will also not discount future consumption much because he is not willing to accept much variability in consumption. If these rates are equal, then holding real consumption constant is utility maximizing.” [P. 40] The authors test allocations and withdrawal rates for retirees with differing levels of secure, non-portfolio-related income. Portfolio depletion results in a “bad” economic state in which consumption depends only on non-portfolio income. “…for each withdrawal rate chosen, the optimal portfolio allocation is the one that is expected to minimize the percentage of bad years…, or variance of consumption for that withdrawal rate.” [P. 43]

**Shortfall risk in consumption v. shortfall risk in wealth**


Insurance carriers are assumed to fail with an annual probability of 0.15 percent. In disasters, the probability of insurance company failure is twice that of government bond default. The discount rate for annuity pricing is based on the yields on ten-year U.S. Treasury bonds. Annuity pricing uses the annuitant population mortality table with an additional load of 10% to cover administration and marketing costs. Survival of retirees, however, is simulated using general population unisex mortality tables. “The success criterion for the strategy search is to minimize the shortfall risk, which is defined as a weighted probability of real income and wealth balances falling below certain thresholds, in a stochastic model. The objective somewhat departs from the conventional analytical assumption that investors maximize their expected utility over consumption and bequests.” [P. 163] Consumption shortfall is defined as a distribution that falls below a pre-established threshold amount. A shortfall in wealth is defined as lacking sufficient funds to cover uninsured contingencies or to leave a targeted level of bequests. In the Pang model, mitigation of consumption shortfall risk competes with mitigation of wealth shortfall risk.

---

43 See the review of the annuity marketplace in Part Four.
The Pang model incorporates the effects of multiple factors including minimum threshold income requirements, investment fees and expenses. The research touches on critical tradeoffs between investment and actuarial approaches to generating adequate lifetime income. Mutual fund investment enables the retiree to improve the budget constraint if returns are good, but exposes the investor to significant consumption declines if returns are poor. Annuity purchases in low interest rate environments also constitute a significant risk. However, delaying an annuity purchase may risk further losses in the mutual fund portfolio with the result that the investor may lack sufficient funds to purchase the desired amount of future annuity income. The risks and benefits of asset management elections are more fully highlighted by such complex and realistic models. The exemplification and quantification of these options within retirement income risk models is fast becoming a prerequisite to effective portfolio management.

**Shortfall risk in terminal wealth v. maximizing lifetime income opportunities**

Closely related to the bequest versus income tradeoff is an assessment of retirement fund strategies in terms of lost opportunity costs. This is a development of the portfolio preferencing criteria that we discussed in the review of the 2004 study by Russell Gerard, Steven Haberman and Elena Vigna [“Optimal Investment Choices Post Retirement in a Defined Contribution Pension Scheme,”]. This line of research views a strategy designed to enhance portfolio sustainability by piling up large amounts of future wealth, absent a bequest motive, as a strategy that diminishes opportunities for lifetime consumption. The issue is, of course, a variation on the lifetime consumption vs. terminal wealth shortfall metrics.

Consider, for example, the 2012 essay by Stephen C. Sexauer, Michael W. Peskin, and Daniel Cassidy [“Making Retirement Income Last a Lifetime”]. The authors point out: “Most of the academic research with respect to retirement strategies has focused on the right tail, where the concern is outliving one’s assets. In our study, we attempted to bring much-needed attention to the left tail, where the concern is getting as much income as possible while a large majority of retirees are still alive.” [P. 77] The authors suggest creating a benchmark to evaluate the success of a retirement income investment strategy. In this case, the benchmark consists of a suitable combination of TIPS and an ALDA—advanced life deferred annuity [see Part Four]. The benchmark is investible—a retiree can implement this two-asset portfolio; and the spendable income which it generates measures how well an investment program designed to ‘beat the benchmark’ is actually performing. Although the authors acknowledge that the best benchmark is an inflation-adjusted immediate annuity, they reject an annuity as a suitable performance evaluation benchmark: “…the cash flows from a real, immediate life annuity are unsuitable as a general benchmark for asset decumulation because the illiquidity of such a strategy is so burdensome that almost no one uses
Their rejection of a SPIA as a performance benchmark, however, does not imply that an SPIA is not a valid measure of the cost of retirement—the current market value of the stochastic present value of retirement cash flow liabilities. The distinction is between a performance evaluation benchmark and a portfolio solvency benchmark. Additional discussion of appropriate retirement performance benchmarks is found in the 2012 essay by Daniel Cassidy, Michael Peskin, Laurence Siegel and Stephen Sexauer [“Be Kind to your Retirement Decumulation Plan—Give it a Benchmark”].

**Conditional vs. Unconditional Shortfall Risk**

Co-authors Michael Finke, Wade D. Pfau and Duncan Williams [“Spending Flexibility and Safe Withdrawal Rates”] develop, in 2012, a model designed to identify solution paths based on expected utility. The model employs a bootstrap methodology, consists of a two asset class portfolio, incorporates the CRRA form of the utility function, and further assumes that the investor has guaranteed non-portfolio income. The authors criticize the use of a shortfall risk metric; but, the guaranteed “pension” income stream allows them to bypass—almost too conveniently—many difficulties that flow from their model’s over-simplified assumptions. The critique of the shortfall risk metric, however, is germane. They assert that shortfall risk analysis may not be the best preferencing criterion for selecting a retirement income strategy: “By emphasizing a portfolio’s ability to withstand a 30- or 40-year retirement, we ignore the fact that at age 65 the probability of either spouse being alive by age 95 is only 18 percent. If we strive for a 90 percent confidence level that the portfolio will provide a constant real income stream for at least 30 years, this means that we are planning for an eventuality that is only likely to occur 1.8 percent of the time. And even that figure assumes that clients are unable to make adjustments to their spending later in retirement. So by relying on standard historical or Monte Carlo simulations to determine a safe withdrawal rate, clients may be unduly sacrificing much of their desired lifestyle early in retirement. The failure to include a client’s willingness to adjust is an important shortfall of the shortfall literature. A common thread in the analysis is that all failures are counted the same, without regard to when the failure occurred or what percentage of the client’s stated aggregate spending goal was funded. Such an all-or-nothing approach to retirement simulation is inconsistent with the way trade-offs are framed in retirement.” [P. 44]

It is not enough for a retiree to know that a shortfall might take place. The magnitude of the shortfall is also of importance—especially if the retiree has resources beyond the financial portfolio. This implies, of course, that any portfolio monitoring and evaluation system incorporating a shortfall risk metric must account for both the probability of the failure as well as its

---

45 We note that a similar, albeit more technical and mathematically elegant critique of unconditional shortfall as a measure of risk, is found in “Shortfall-Risks of Stocks in the Long Run,” Peter Albrecht, Raimond Maurer & Ulla Ruckpaul. *Financial Markets and Portfolio Management* vol. 15 no. 4 (2001), pp. 481 – 499. This article, however, is not often cited in the literature because it deals with data from the German stock and bond markets.
magnitude should the failure occur. Interestingly, for many values of the CRRA risk aversion coefficient, their model identifies an optimal withdrawal rate / asset allocation strategy that does not minimize shortfall risk. In fact, the investor elects portfolio management strategies, under a utility-based preferencing metric, that have a relatively high likelihood of depleting the portfolio during the life of the retired investor.

**Shortfall Risk and high risk aversion v. Fisher utility and high subjective discounting**

Moshe Milevsky’s 2012 book *[The 7 Most Important Equations for Your Retirement]* discusses Irving Fisher’s viewpoint on the utility of retirement consumption. Determination of the optimal consumption level includes factors for (1) investor risk aversion, (2) subjective discounting of the utility of future consumption (e.g., a preference for income in early retirement assuming good health and the capacity to enjoy leisure activities), and (3) longevity expectations. In terms of setting a retirement income budget, the risk aversion factor reflects a concern with outliving financial resources. A high co-efficient of risk aversion leads to a conservative spending rate in early retirement. Conversely, a high subjective preference rate discounts the value of future income and leads to higher spending during early retirement. This tug of war is, of course, one aspect of the general risk/reward tradeoff faced by retired investors.

Milevsky states: “Irving Fisher the economist was the first to properly formulate how *rational consumers* should adjust their consumption spending over time. This is the intertemporal aspect of economic tradeoffs.” [P. 78] Data from the US Department of Labor suggests that “by 65, retirees are spending between 50% and 70% of what they did at 50. And, by 80 it has dropped to under 60%.” [P. 83] Milevsky quotes from Fisher’s 1930 study: “Uncertainty of human life increases the rate of preference for present over future income for many people….“ [P. 91] Investors exhibiting Fisher utility prefer to spend and enjoy scarce resources today rather than husband them against a remote contingent probability that they might be needed at an advanced age. Milevsky also notes that Yaari’s 1965 article translates Fisher’s view of investor utility into a mathematical expression suitable for Life-Cycle modeling.

The Fisher view of retirement income utility has several implications. Seemingly, it stands in an uncomfortable relationship to a ‘budgetary-certainty’ approach. Many retirees derive satisfaction from knowing that the portfolio can produce a steady monthly constant-dollar income. Indeed, throughout much of this literature review, we comment on retirement income risk models that emphasize achievement of a sustainable *threshold* income in each period. The threshold represents the lower-bound income that is acceptable to the retiree. The budget goal is akin to the “hurdle race problem” discussed in the previously cited 2003 paper by S. Vanduffel, J. Dhaene, M. Goovaerts, and R. Kaas. Some models

46 Or, a low elasticity of intertemporal substitution.

47 Of course, any budget must provide funds at a minimum ‘subsistence’ level.
assume a HARA or a state-preference utility function the violation of which, given the retired investor’s preferences and constraints, generates disutility. In these models, maximization of additive utility across all states gives way to achieving, at least, a positive “utility hurdle” in each state. Furthermore, the introduction of an income threshold creates the conditions under which calculation of the free-boundary—the portfolio’s feasibility condition—is both necessary and possible. This, in turn promotes the creation of risk models that insure compatibility between a utility-based portfolio preferencing system and a system based on safety-first or shortfall risk metrics.

Additionally, we note a predilection by some retired investors to back-load their retirement income spending policy. This might, for example, be the case where, in previous generations, family members faced significant end-of-life costs; or, where there is a significant age difference between spouses. Such a spending preference turns the Fisher optimal retirement consumption path on its head. It is further proof that pre-set rules, combination of rules, and one-size-fits-all bright line spending policies are problematic.

The Gordon Pye Model: Retirement Planning with Scarce Resources

In 2012, former U.C., Berkeley finance professor Gordon Pye published a book [The Retrenchment Rule] on retirement income strategies for investors owning modest-sized portfolios. The book’s premise is that many, if not most, retired investors lack a sufficient amount of capital to finance sustainable lifetime income at a level that can permanently preserve their prior standard of living. Additionally, many do not have the option to continue in their jobs past normal retirement age. Given the economic reality which these retirees face, “…the key question in spending retirement assets [is] not achieving sustainable withdrawals, but when to retrench.” [P. xv] There is no doubt that spending cuts must occur. However, if the initial cut in spending is so great that it forces an unacceptably high level of immediate economic pain, then a more gradual schedule of reductions may be the preferred alternative provided that the reductions do not result in unacceptably low future withdrawals.

Pye advances the following argument: “The initial withdrawal should be the one required to provide the prior or desired standard of living subject to a limit. That limit should be the largest withdrawal that could be made without increasing too much the risk of low withdrawals later in retirement. Moreover, this limit should be based on the discount rate that gives the best results in a series of simulated withdrawals.” [P. xv] The discount rate that gives the best results—i.e., balancing sufficient current income against the probability-adjusted likelihood of lower future cash flow—is the Retrenchment Discount Rate, the application of which is the ‘Retrenchment Rule.’ Pye’s retirement withdrawal strategy is aimed at retirees who exhibit a strong preference to maintain their current standard of living [Habit Utility] but who own resources insufficient to support the required cash flows throughout long
planning horizons: “…the present value of the withdrawals required to sustain their existing standard of living exceeds the value of their initial investment. Thus, these retirees will have to retrench.” [P. 269]

Classic Life-Cycle utility-based models tell us that the investor seeks to optimize utility by following a smoothed consumption path. A drastic reduction in spending at the moment of retirement, however, creates an unacceptable ‘discontinuity’ in the consumption path.

Given the book’s premise, following the four percent initial withdrawal strategy will not work for many retirees because “…many individuals have to begin retirement without having saved nearly enough to cover essential expenses with a 4 percent withdrawal.” [P. 1] Although the 4 percent rule often forces draconian budget reductions, Pye notes that such reductions may prove unnecessary if (a) initial investment returns are above expectation; or (b) the investor suffers a health decline that adversely effects longevity expectations.48

Pye illustrates the difficulties with the 4 percent rule. He conducts two tests assuming an annually rebalanced portfolio of 75% S&P 500 and 25% intermediate US government bonds. The first test applies the 4 percent rule to an age 65 investor, in good health, beginning retirement in 1991. Despite the bear markets in stocks both early and late in the first decade of the 21st century, the retiree is able to sustain a 7.5 percent withdrawal rate through 2010—the time of the book’s composition. However, if the hypothetical retirement starts in 1966, an initial withdrawal strategy of 7.5 percent would have required gradual reductions to avoid portfolio depletion. By 1982, the retiree can withdraw only 2.1 percent of the initial portfolio value. As it turns out, 1966 was the worst year to retire of all years since 1926. By contrast, 1991 was the most favorable year for retirement. Pye’s point is this: following the 4 percent rule would result in a significant and unnecessary permanent reduction in income for the 1991 retirees. For 1966 retirees, the 4 percent rule would have curtailed early retirement income substantially [7.5% v. 4.0%] and would have fully depleted the portfolio by 1996 [30 years]. “Those who withdrew 7.5 percent in 1966 eventually had to retrench somewhat more. But they had 10 years [before they hit the 4 percent withdrawal target level] over which to plan and make these adjustments….Also, with good health, and perhaps some pent-up enthusiasm for activities such as travel, increments of spending early in retirement are likely to provide more satisfaction than equal increments later on.” [P. 6] Bottom line—the 4 percent rule mandates an immediate and substantial reduction in many retirees’ standard of living. Drawing on observations similar to those made by Irving Fisher,49 Pye states: “The Retrenchment Rule weighs whether this immediate pain is justified given that the future funds provided may turn out to never be

---

49 Fisher is not mentioned in the book.
needed.” [P. 1] The Retrenchment Rule requires a more modest type of ‘glide path’ reduction and, of crucial important, this reduction is implemented only if an unfavorable sequence of returns unfolds.

Clearly, a primary control variable for Pye’s version of dynamic asset management is the withdrawal amount. Examination of all historical periods since 1926 reveals that withdrawals beginning at the 7.5 percent rate stay above the 4 percent rate in 70 percent of the cases. Cuts should be made when the withdrawal rate is unsustainable. But retirees do not know \textit{ex ante} the future sequence of returns, inflation rates, or health changes which they will face. Therefore, according to Pye, the utility maximizing retiree avoids making immediate and economically painful cuts when such cuts may, in fact, prove unnecessary. Any retiree with a positive time preference for consumption will wish to avoid painful retrenchment forced upon them at the beginning of their retirement by a four percent withdrawal rule.

Pye’s argument for beginning initial consumption in retirement at a relatively high amount seems to run counter to recommendations made by commentators using a shortfall or risk-of-ruin portfolio monitoring metric. Such, however, is not the case. Pye’s argument is more along the lines of: cut back when necessary; cut back as gradually as possible to preserve a smoothed consumption path; and make sure that the portfolio does not outlive the investor. Fundamentally, Pye asks the investor to make deliberate, well-considered asset management elections rather than defaulting to a fixed, auto-pilot withdrawal rule designed to protect the portfolio against low-probability events. As such, Pye’s suggestions conform to both the classical utility-based approach to retirement portfolio management as well as to the shortfall avoidance approach.

The following paragraph, for example, reveals Pye’s academic grounding in Life-Cycle utility modeling: “Future investment withdrawals of the same real value are worth less the further in the future they are expected to occur. One reason is that it becomes increasingly less likely that retirees will survive that long. Another is that many havepent-up plans for activities after they retire such as travel, but these desires become satisfied. Also, later on withdrawals of the same real value are likely to provide less satisfaction as lifestyles slow and mobility declines even for those who remain in as good health as can be expected. To reflect this decreasing value the Retrenchment Rule discounts future withdrawals by a constant rate of interest for each year in the future until they will be made. The value of this discount rate is selected by simulating the use of the rule with different rates over a hypothetical retirement period. This is to see which rate gives the best performance. The discount rate selected is called the Retrenchment Discount Rate, or RDR.” [P. 12]
However, Pye also recognizes that retired investors are averse to the risk of outliving financial resources. Although Pye does not use the term ‘feasibility condition’ when discussing decumulation strategies, he advances an argument that is compatible with this concept: “To reflect their declining value future withdrawals are discounted by a constant rate of interest for each year in the future until they occur. This discounting gives the present value of a future withdrawal based on the discount rate that has been used. Adding up these discounted values for each of the withdrawals gives the present value of the stream. This is the present value of the future withdrawals in real terms needed to provide the existing standard of living. The present value of the funds available to make these withdrawals is the current value of the investment portfolio. Suppose the present value of the stream of withdrawals required to provide the existing standard of living is less than the value of the portfolio. Then retrenchment is not required….there are sufficient funds to provide the existing standard of living in the future. On the other hand, suppose that the present value of the stream to provide the existing standard of living exceeds the value of the portfolio. Then retrenchment is required.” [P. 21] This language puts Pye close to the school of retirement portfolio management that focuses primarily on period-by-period income sustainability; and, it conforms to a free-boundary / portfolio solvency monitoring approach.

A key issue is how to determine the appropriate discount rate. The best discount rate is the RDR; and this rate is determined by calculating the present value of the stream of future withdrawals required to provide a given standard of living—the desired discount rate—limited, in turn, by the discount rate that gives the ‘best’ results—where best is defined as the most appropriate tradeoff between current cash flows and possible lower future cash flows—across the distribution of simulated results. Simulations assume a maximum limit on life expectancy of age 110.

For example, suppose that a retired investor in good health requires an initial withdrawal rate of 8% from their modest-sized portfolio. At this rate, she can avoid making a painful cut in her standard of living. If the RDR with the best tradeoff between current cash flow adequacy and the likelihood of future painful income reductions is 6%, however, then some immediate retrenchment must occur. “Suppose that the value of the investment is 100 so that the required withdrawal is 8. Input n=45 [110-65=45], i=6.0,

50 Pye’s discount rate approach is, in some respects, the opposite side of the coin from the “equivalent payment value” developed by Hughen, J. Christopher, Laatsch, Francis E. and Klein, Daniel P. 2002. “Withdrawal Patterns and Rebalancing Costs for Taxable Portfolios.” Financial Services Review, Vol. 11: 341-366. The equivalent payment value expresses terminal wealth in terms of the extra periodic payment that could have been received throughout the planning horizon. It is “…calculated using an interest rate equal to the total return on equity over the particular time period.” [P.363] This value is then expressed as a percentage of the initial portfolio value. Assume $1 million portfolio with terminal wealth of $1.4 million. Total annual return on equity for a 20-year period is 10%. If this had been converted into a nominal annual payment the amount of extra yearly income would be $24,443. As a percentage of initial portfolio value, terminal wealth was sufficient to support a 2.4% increase in the nominal withdrawal rate. Given that the equivalent payment value is a backward looking calculation, it has limited use in a portfolio monitoring system.
PMT=8, and FV=0. Calculate the present value of this stream getting 131.1, which exceeds the value of the investment of 100. When the withdrawals have declined from 8 to 6.1 their present value has fallen to 100. As this is the value of the portfolio this withdrawal is allowed. It is the largest withdrawal allowed by the Retrenchment Rule with a discount rate of 0.6.” [P. 25] Another way of looking at this calculation is that “It is the largest fixed annuity that can be obtained each year from the investment for the longest that the retiree might live. This is when the investment earns a return equal to the assumed discount rate of .06. A larger stream of withdrawals than this annuity will have a present value that exceeds the value of the investment and will require retrenchment.” [P. 25] It is worth noting that the discount rate [RDR] for this hypothetical annuity is not derived from the current term structure of interest rates. Neither is it the expected portfolio rate of return. Rather, it is a discount rate derived from a trial and error process over the entire distribution of simulated portfolio returns.

Here’s how Pye’s example plays out in a dynamic portfolio monitoring and withdrawal setting: “For making the simulations it is assumed that 6.1 is withdrawn at the beginning of the year and spent over the year. Another withdrawal is not made until the beginning of the following year. Suppose the real return on the portfolio over the coming year is .03. The value of the portfolio at the beginning of the following year before a withdrawal is made is then (100-6.1)(1.03), or 96.72. This is in real terms because the return of .03 is in real terms. The 96.72 is 96.72 percent of the initial investment because the initial investment is equal to 100. The Retrenchment Rule is now applied just as it was initially to get the next withdrawal. All of the inputs have changed, however, except that the discount rate is still .06. To calculate the largest allowed withdrawal input n=44, PV=96.72, i=6.0, and FV=0. Calculate PMT getting 5.9 as the largest withdrawal allowed by the Retrenchment Rule. As 6.1 must be withdrawn to sustain the prior standard of living some retrenchment is required. Suppose that a severe bear market occurs over the coming year and that the real return is -.25 instead of .03. In this case the value of the investment at the beginning of the following year is (100-6.1)(.75), or 70.42. Calculating PMT in this case the largest allowed withdrawal is 4.3. Now a major retrenchment is required. The withdrawal must be reduced from 6.1 to 4.3. Suppose only 4.0 had been withdrawn initially instead of 6.1. In this case a major retrenchment is required initially, but no retrenchment is necessary at the beginning of the following year. The value of the investment is now (100-4.0)(.75), or 72.00. But this is only slightly higher than the 70.42 obtained with the 6.1 withdrawal. Thus, suppose a severe bear market occurs in the coming year. It then makes little difference if the Retrenchment Rule is used with a .06 discount rate or there is a 4 percent withdrawal.….The advantage of the higher withdrawal is the very strong chance that a much better return will be earned over the coming year. In this event major retrenchment may never be necessary….If only 4 percent is withdrawn initially a major retrenchment occurs for sure immediately.” [P. 26]
Portfolio management elections are not based on an expected future return, but rather on current observables—age and portfolio value. The extreme conservatism of calculating the applicable planning horizon based on the maximum age in the mortality table balances the extreme liberalism of higher initial withdrawal rates. This is a key point of differentiation between Pye’s portfolio monitoring and evaluation method and others. It is almost as if Pye is betting that the two ‘error terms’—longevity span and withdrawal amounts in excess of what can be sustained under all historical conditions—will cancel, thus leaving the investor with the expectation of an adequate retirement income. By contrast, alternative monitoring and portfolio evaluation approaches, using an annuity benchmark, directly incorporate the force of mortality and the term structure of interest rates via the annuity pricing factor.

The process of recalculating the allowable withdrawal amount continues each year through a maximum age of 109. This includes simulation of the distribution of future returns and recalculation of the most appropriate discount rate. “Suppose next that the discount rate is .08 instead of .06. In this case the largest allowed initial withdrawal is the value of the annuity obtained when the investment earns a return of .08 instead of .06. The largest allowed initial withdrawal in this case rises from 6.1 to 7.6 percent. If the required withdrawal to avoid initial retrenchment is still 8 percent the initial withdrawal increases to 7.6 percent.” [P. 28] This is the limit imposed by the annuity calculation for n=45 and FV=0. If a major health crisis occurs, planning over the long horizon may be sufficiently conservative so that funds will be available to provide the needed liquidity. Thus, Pye views the lack of precision in his monitoring system as a virtue in that ‘overstating’ the probability of a long lifespan is equivalent to building in a reserve in the event that expected longevity is cut short by health changes.

The gist of Pye’s observations is that a higher discount rate [RDR] provides higher withdrawals early in retirement. However, the high withdrawals may deplete the portfolio more rapidly than a withdrawal strategy calculated under a lower RDR. Selection of a high RDR results in a higher probability of future retirement retrenchment. Assuming a normal return distribution with stocks having an expected return of .07 and a standard deviation of .18, all retirees will select a discount rate of .08 because the expected withdrawals at this rate are higher than those for any lower RDR. Thus, .08 exhibits a property of dominance at each age. For example, a discount rate of .10 offers a higher probability of a lower withdrawal at age 90 than does an RDR of .08. However, a retiree may still select an RDR of .10 if his or her time preference factor (‘impatience’) is high. If, however, either of the two parameters estimates is inaccurate—e.g., lower mean—then .08 may not be the optimal RDR. Pye suggests that .08 is probably not too far from the optimal value; and, therefore, investors may still prefer to stick with it because they will be adequately compensated by the opportunity to avoid severe retrenchment in early retirement.
Pye stresses the importance of the subjective time preferencing / discount rate. He notes that a key objective of retirees “…is to avoid low withdrawals at any time over their retirement. But retirees are likely to be more averse to low withdrawals early in retirement than later on….If retirees are less averse to low withdrawals later in retirement this difference should be reflected in the utility function. To do so later withdrawals can be multiplied by an equivalence factor to make their effect on utility comparable to that of earlier withdrawals.” [P. 271] This means that the retirement income goal is to maximize time-adjusted utility. For a simple utility function, each future time-adjusted withdrawal generates utility equal to that of the initial withdrawal divided by the time equivalence factor for the applicable future year. Retirees wishing to spend more in the early years of retirement will have a larger subjective time preference rate. When the preferencing rate is equal to zero, the discounting is simply $1 +$ risk free rate. This is characteristic of a retiree that prefers a constant standard of living. Often, however, retirees “…prefer some downward slope in their withdrawals.” [P. 284]

Pye’s modeling is based on a sophisticated treatment of input variables. He develops a two-state model based on market valuation measures (e.g., price/earnings ratio, and dividend yield) and, when the metrics evidence an abnormal valuation level, expected return is adjusted accordingly. However, Pye does not fall into the trap of market timing. Although it may be tempting to change the stock/bond weighting because of changes in relative valuation, this is often counterproductive: “This is not surprising as changes in valuation account for only a very small portion of changes in realized investment return.” [P. 205] The model treats inflation as a stochastic variable and accounts for investment fees and, where appropriate, tax liabilities.

In contrast to Pye, a 2014 working paper by Patrick J. Collins Huy Lam and Josh Stampfli [“Monitoring and Managing A Retirement Income Portfolio”] presents a retirement income risk model that accommodates customized patterns of retirement spending including constant real consumption, front-loaded, and back-loaded consumption. Assume, for example, that a retired investor’s wealth is greater than the present value of periodic minimum threshold needs as determined by reference to a SPIA benchmark—i.e., the free boundary. Collins, Lam & Stampfli modify a Fisher utility-based model by concurrently incorporating a state-preference utility function. In this configuration, optimal spending is a combination of:

1. A minimum floor income requiring a fixed constant dollar distribution; and,
2. A floating distribution component based on investor time preference rates—e.g., the rate of consumption declines as age advances—and on spending constraints conforming to investor risk aversion to outliving financial assets, as well as investor preferences for gifting and bequests.
The free boundary conditions allow the investment advisor to recommend asset management elections based on an intelligent assessment of investment “surplus” in the face of dynamic consumption and terminal wealth preferences.

**Conclusion**

Advances in retirement income risk modeling are striking both in terms of the complexity of the models and the scope of insights engendered. However, many of the case examples offered in the literature do not comport comfortably with likely spending patterns faced in retirement. Retired investors rarely spend according to constraints established either by shortfall probability estimates or according to auto-pilot formulae like the 4% adjusted-for-inflation rule. Furthermore, the utility-based analysis underpinning many Life-Cycle models generates optimal consumption rules based on the form of a possibly linear utility function rather than on the practical choices and exigencies which the investor encounters. Financial planning recommendations flowing from such risk models appear to carry a significant import; however, investors should be mindful that such recommendations often arise in highly artificial contexts.

For example, consider a spending pattern that an investor on the threshold of retirement might wish to test in terms a portfolio’s ability to fund it adequately:

1. During each month of retirement, withdraw 1% of the average value of the portfolio calculated over the previous 36 months;
2. For the first 63 months (e.g., the time remaining on a mortgage obligation), distribute an additional $8,000 per month adjusted for inflation;
3. In months 64 through 180 reduce the additional distribution to $6,000 per month adjusted for inflation;
4. Thereafter, distribute a constant-dollar $4,000 per month decreasing at a 2% per year rate.

This pattern exhibits a consumption tilt towards the early stages of retirement. It combines a high degree of budgetary control with flexibility to reap the rewards of potential future portfolio increases. The distribution suggests that the investor is willing to decrease later-life consumption if he or she survives longer than 15 years. The late-in-life strategy might assume that the targeted percentage reduction in fixed-amount portfolio distributions may, to some extent, be offset by CPI-increases in the investor’s Social Security benefits; and, if investment results are satisfactory by the 1% “profit sharing” piece of the withdrawal strategy. The investor may also value additional customization of spending policy to test the economic consequences of making lifetime charitable or intrafamily gifts. As the investor decides how to implement retirement—when to retire, whether to work part-time, how to adjust spending, etc.—there can be little doubt that the attention is directed more towards finding out if the portfolio can sustain the
preferred cash-flow pattern as opposed to, say, the historical success rate of the 4% withdrawal rate rule. Optimal retirement planning puts the investor in a position to reveal utility by selecting among various retirement spending patterns and terminal wealth outcomes. This process differs from budgets imposed by pre-determined rules.

Even the sample spending pattern, however, is too neat. The retired investor faces a possibility of incurring unexpected financial emergencies that will force liquidation of financial assets. If dental, home repairs or other non-discretionary outlays suddenly require unanticipated portfolio withdrawals, the expenses must be paid irrespective of whether they are accounted for in the annual budgeted withdrawal amount. If initial resources are modest, the investor often wrestles with the tradeoff between the ability to sustain an aspirational standard of living and the possibility of future financial hardship. It is useful, as the literature survey makes patently clear, to distinguish between an aspirational standard of living and a threshold or minimally acceptable standard of living. Although retirees seek to preserve the aspirational standard of living for as long as possible, they are constrained by the feasibility condition. That is to say, it is imprudent to engage in a level of current spending that creates a high probability of failing to meet threshold future living costs. If we cannot say that a commercial annuity constitutes the optimal retirement income path, we can infer that the prudent investor checks the portfolio’s financial health against an annuity-based benchmark. The next section explores the nature and scope of such a benchmark in greater detail.
PART THREE: A SURVEY OF ACADEMIC LITERATURE ON ANNUITIES

Estimating Annuity Costs and Loads & Utility

Value per Premium Dollar, Expected Present Discounted Value, Internal Rate of Return, Implied Longevity Yield, the Money’s Worth Ratio, and Annuity Equivalent Wealth

Previous sections of the literature review discuss (1) the merits of using a Single Premium Immediate Annuity [SPIA] as a monitoring and performance benchmark, and (2) the merits of electing to annuitize some or all of financial wealth to guarantee lifetime periodic income payments. The latter asset management election is often evaluated in a Life-Cycle model seeking to gauge the utility value of an actuarially fair annuity. Actuarially fair annuities are, however, unavailable to investors. The prudence of exercising an option to annuitize depends, of course, on a variety of factors including contract costs. The cost of an actuarial solution determines the capital sacrifice required to transfer longevity risk from the investor to the insurance industry. That is to say, it quantifies the amount of wealth that must leave the financial asset portion of the retirement portfolio in order to secure a target amount of periodic lifetime income.

There is a large body of research on the topic of annuity costs and benefits. Unless otherwise stated, this literature review focuses on SPIAs promising either fixed or inflation-adjusted payouts. The review begins in 1998 when Moshe Milevsky [“Optimal Asset Allocation Towards the End of the Live Cycle: To Annuitize or Not to Annuitize”] uses a Value per Premium Dollar [VPD] method to estimate an annuity’s load. An actuarially fair annuity has a VPD equal to one; a commercial annuity has a VPD less than one. The calculation formula is a two-step process. Step one sums, over a maximum possible lifetime (through age 115), the values of a fraction where the numerator is a $1.00 periodic annuity payment adjusted for the probability that the annuitant is alive to receive it, and the denominator is the applicable time-value-of-money discount rate. Step two evaluates the ratio of the sum of the discounted, probability-adjusted $1.00 payments (step one) to the current market price of an annuity contract. Milevsky estimates the average annuity load for Canadian annuities during 1984-1996 is approximately 12% when discounting by the corporate bond rate.

In 1999, Mitchell, Poterba, Warshawsky and Brown [“New Evidence on the Money’s Worth of Individual Annuities”] outline three methods to determine annuity costs and value.

1. The first method, which they term the expected present discounted value [EPDV] of an annuity, is the step one calculation of the VPD—the sum of periodic mortality-adjusted payments
discounted by an appropriate term structure of interest rates. The authors estimate implies annuity costs of between 7% and 16%.

2. The second method calculates an annuity’s internal rate of return (IRR). The IRR formula subtracts the commercial annuity price from the actuarially fair EPDV, and solves for the EPDV discount rate that brings the cost difference to zero. The investor can determine the relative attractiveness of annuitizing some or all of current wealth by comparing the annuity IRR to other investments.

3. The third method of estimating annuity cost and value compares the expected utility gained by annuitization to the expected utility provided by alternative investments. In this case, the authors calculate the percentage of wealth that must be annuitized in order to produce an equal utility value of consumption absent annuitization. Specifically, the utility of financing consumption through risky investment is equivalent to a wealth reduction of between 30% and 38% when compared to financing consumption through an actuarially fair annuity for an age 65 investor exhibiting constant relative risk aversion [CRRA].

Thus, assuming no pre-existing annuity income, a CRRA investor could, in the authors’ opinion, increase utility by annuitizing wealth even when facing loads as high as 30%.

In 2000, Estelle James and Dimitri Vittas [“Annuity Markets in Comparative Perspective”] define an annuity’s money’s worth ratio [MWR] as the present value of the expected stream of benefits divided by its initial cost; or, in terms of the previous cost measures, the EPDV ÷ Market Price of a Commercial Annuity. James and Vittas find that the MWR in many countries approaches a value of one when discounted at the risk-free rate. Inflation-adjusted annuities, however, have ratio values 7% to 9% lower than comparable nominal payout annuities. The comparatively high MWR ratio values found by James and Vittas are, in large part, a byproduct of their choice of the discount rate. Future studies often utilize multiple discount rates—treasuries, corporate bonds, mortgage-backed securities, and so forth. The MWR ratio is highly sensitive to the discount rate choice and, therefore, care must be taken to understand how an economist or actuary calculates MWR. For example, the choice of the discount rate can reflect either the predominant components of insurance company investment portfolios (discounting from the seller’s perspective), or reflect the rates of return available to annuity buyers (discounting from the buyer’s perspective). N. Charupat, M. Kamstra and M.A. Milevsky, in their 2012 essay “The Annuity Duration Puzzle,” note that “The 10-year swap rates are the best match to the average duration of the annuities we look at….“[P. 7] In calculating Annuity Equivalent Wealth, Michael J. Zwecher argues that annuity payments should be discounted at the cost of debt financing for the issuing carrier: “…insurance contracts for retirees are little more than corporate debt with a mortality pooling component; the correct
rate to discount payments that an insurance contract will pay would be at the rate that the insurance company would pay to issue debt at that maturity….“[P. 56]

Several research papers reconsider the assumptions underlying calculation of the MWR numerator. Selection of mortality data—e.g., general population mortality vs. annuitant subpopulation mortality—significantly changes MRW values. Olivia S. Mitchell emphasizes this point in 2001 ["Developments in Decumulation: The Role of Annuity Products in Financing Retirement"]: “Mortality processes may be heterogeneous across subgroups of the population.” [P. 10] Mitchell’s essay develops the concept of annuity equivalent wealth [AEW] which first appears in the previously cited 1999 essay co-authored with Poterba, Warshawsky & Brown. Mitchell defines AEW as the additional amount of financial wealth required to bring the utility of consumption financed from annuitization into equilibrium with the utility of consumption financed from a risky asset investment portfolio. An annuity contract guarantees a specified consumption level; a risky asset portfolio may provide funds sufficient to meet future consumption needs. The AEW of an annuity for a highly risk averse investor may be significantly positive despite the fact that the MWR value is less than 1. She considers AEW under various inflation processes for consumers with no pre-existing annuity benefits as well as for consumers who have previously annuitized half of their wealth—e.g., by claiming Social Security benefits. The inflation process confronting the investor has a significant impact on AEW especially if the annuity contract provides only nominal periodic payments. As a general rule, however, Mitchell finds, depending on the investor’s coefficient of risk aversion under a CRRA model, that financial wealth earmarked for production of income must be adjusted upwards by 30% to more than 200% to provide utility values equal to annuitization. Risk-averse CRRA investors consider longevity insurance to be a valuable benefit.

The concept of AEW is further developed in the 2001 essay co-authored by Brown, Mitchell and Poterba ["The Role of Real Annuities and Indexed Bonds in an Individual Accounts Retirement Program"]. The authors are primarily interested in determining the extent to which AEW can “overcome” an EPDV value less than 1, where AEW is defined as “…the amount of wealth that a retiree needs—if he does not have access to an annuity market—to achieve the lifetime utility level that he can attain with access to an annuity market.” [P. 346] Despite the fact that stocks offer a higher expected payout, they expose investors to market volatility, and cannot guarantee a fixed real return. An alternative, under the authors’ model, is a constant-dollar annuity. The AEW values for such a hypothetical, fair-valued annuity range from 1.502 to 2.004 for coefficient of relative risk aversion ranging from 1 through 10. The model assumes an investor with no pre-existing annuity income. Although some commercial annuity contracts provide a constant-dollar income stream, such an income stream may not be optimal. When an investor fails to exhibit a CRRA utility function or when the investor’s personal (“subjective”) discount rate is
high because consumption early in retirement is more highly valued than consumption towards the end of
life (“impatience”), an annuity income may not prove attractive. Modeling flexibility becomes
increasingly important in future research studies as retirement risk models begin to incorporate different
assumptions regarding the structure of the utility function, the nature of the return generating process, and
the number of exogenous independent variables. The authors review the annuity market in the United
Kingdom and estimate that the EPDV of a nominal UK annuity is approximately 90% of the premium.

A further extension of the concept of AEW appears in J. R. Brown’s 2001 research paper [“Private
Pensions, Mortality Risk, and the Decision to Annuitize”] which employs dynamic programming to
construct an AEW measure consistent with CRRA utility. Brown’s model suggests that wealthier
individuals are less likely to annuitize. Among the possible explanations are:

1. Less likely to exhaust financial resources;
2. Have more pre-existing annuity wealth;
3. More likely to try to earn high investment returns; or,
4. More likely to have a bequest motive.

A 2002 study by Brown, Mitchell and Poterba [“Mortality Risk, Inflation Risk, and Annuity Products”]
finds that the EPDV of a US annuity; is between 80 to 90 cents per dollar of purchase price for members
of the general population; but between 90 and 100 cents per dollar for members of the annuitant-mortality
table population. Finally, S. Browne, M.A. Milevsky and T.S. Salisbury [“Asset Allocation and the
Liquidity Premium for Illiquid Annuities”] question, in their 2003 study, the extent of an actuarially fair
annuity’s utility value given capital market demand for increased returns on illiquid (hard to trade) assets.

called Implied Longevity Yield [ILY]. He notes, assuming constant annuity pricing factors, that the cost
of providing a dollar of lifetime income at older ages is less than the cost at younger ages. The ILY is the
investment return needed to withdraw an annuity cash flow measured at age ‘x’ so that the investor has a
sufficient portfolio value to purchase, at an older age ‘y’, an equal or greater lifetime annuity cash flow.
Of course, the risk of the strategy is either (1) a lower-than-anticipated investment return fails to produce
sufficient wealth to purchase an equivalent income stream, or (2) that the future annuity cost increases
because of changes in market interest rates or longevity expectations. Any self-annuitization plan must
produce an earnings rate at least $\lambda$ above the insurer’s annuity pricing rate in order to make the self-
annuitization plan reasonable. The investor must beat the insurer’s pricing benchmark—(e.g., cost of
capital, bond portfolio return, etc.) plus the annuity’s mortality credits ($\lambda$). Solving for the ILY provides
insight into the feasibility of a successful deferral—how likely is it that the investor can earn the ILY or
better at a level of risk that is comfortable? Although Milevsky adopts an option valuation approach to investment decision making, the concept of ILY captures some of the surveillance and monitoring issues faced by retired investors. Milevsky’s discussion complements the previously cited analysis of the distribution of future annuity costs undertaken by Feng Li. Investors may face a critical asset management election: as the value of the portfolio draws towards the free boundary should they continue to own a risky asset portfolio? Quantifying the yield required to sustain a self-annuitization investment strategy is a useful first step in prudent decision making.

Recent academic research into the value of annuitization usually employs a utility-based framework. The 2012 article by Nathan Zahm and John Ameriks [“Estimating internal rates of return on income annuities”] is a noteworthy exception. They define IRR as “the rate the annuity payments are discounted to equate them to the annuity purchase price.” [P. 2] It is difficult, in the authors’ opinion, for consumers to know the cost of an annuity contract: “Purchasers simply see an all-in annuity quote as a single ‘net yield’ offering and must assess the attractiveness of the annuity arrangement on that all-in basis….​” [P. 5] An annuity load consists of four elements: (1) conservative pricing reflecting anti-selection risk to the insurer; (2) cost of maintaining a reserve against the risk that the annuitant population may realize greater-than-expected mortality improvements; (3) administrative costs; and, (4) profit. “Costs arising from adverse selection in the insurance market and from administering the annuities are substantial.” [P. 9] The IRR evaluation metric helps consumers determine if annuitization is an attractive retirement income strategy. Given annuity contract pricing in year 2012, the article calculates IRR for both nominal and inflation-adjusted contracts issued at various ages. The IRRs are based on the median life expectancy and are based on either 10 or 20 year US Treasuries or, for CPI-linked payouts, on 10 or 20 year TIPS. The 10 year securities are used for the 75 and 85 year old purchaser; the 20 year security for the 65 and 70 year-old.

The table of nominal IRRs is as follows:

<table>
<thead>
<tr>
<th>Age</th>
<th>Treasury Rate</th>
<th>Male IRR</th>
<th>Female IRR</th>
<th>Joint IRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>3.02</td>
<td>3.27</td>
<td>3.27</td>
<td>3.18</td>
</tr>
<tr>
<td>70</td>
<td>3.02</td>
<td>2.37</td>
<td>2.69</td>
<td>2.74</td>
</tr>
<tr>
<td>75</td>
<td>2.28</td>
<td>0.52</td>
<td>1.81</td>
<td>2.04</td>
</tr>
<tr>
<td>80</td>
<td>2.28</td>
<td>-4.03</td>
<td>-0.55</td>
<td>0.45</td>
</tr>
</tbody>
</table>

The table of inflation-adjusted IRRs is as follows:
The authors also calculate IRRs for the top quartile of the population. These are the individuals who end up living longer than 75% of the annuitant population and, therefore, realize a better payoff. The authors conclude: “The internal rates of return for this group are all above current yields available on investments with comparable investment risk over similar horizons.” [P. 7]

Jeffrey Brown revisits earlier research in a 2011 paper [“Longevity-Insured Retirement Distributions: Basic Theories and Institutions”]. The updated study concludes that annuities are not attractively priced for the general population: “First, insurance companies selling annuities need to cover administrative and marketing expenses and earn a competitive accounting profit. Second, to the extent that individuals who choose to annuitize have longer life expectancies than the general population, insurance companies need to adjust their prices to reflect this fact.” [P. 61] The study estimates that “…administrative costs account for a 3 to 5 percent reduction in annuity payouts….adverse selection is responsible for an 8 to 12 percent reduction in annuity payouts.” [P. 62] The article identifies a list of factors that may serve to reduce the value of a nominal annuity. For example, the welfare gains from annuitization are not as great for a married couple as they are for individual annuitants. Additionally, nominal annuities are particularly vulnerable to persistent inflation. Brown calculates that the purchase of a nominal joint and 50% survivor annuity in a 3.2% annual inflation environment by a household with a 65-year-old man and 62-year-old woman who have 50% of wealth pre-annuitized generates an AEW of 0.88. “…that is, the couple faces a 12 percent load factor on their annuity purchase.” [P. 66]

Any utility based measure of annuitization is, of course, model dependent. The classic Life-Cycle models of consumption smoothing—savings during working years and decumulation of financial assets during retirement—generally solve for utility optimization. As we cautioned earlier, readers of academic papers should be aware of the implications of the choice of utility function for the retirement income model. Guan Gong and Anthony Webb note in their 2012 paper [“Evaluating the Advanced Life Deferred Annuity—An annuity people might actually buy”] that their model’s outputs on the AEV of both inflation-adjusted and nominal annuities are highly sensitive to the use of CRRA utility: “The above calculations are contingent on a utility function that does not appear to be very predictive of current

<table>
<thead>
<tr>
<th>Age</th>
<th>Treasury Rate</th>
<th>Male IRR</th>
<th>Female IRR</th>
<th>Joint IRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>0.74</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.07</td>
</tr>
<tr>
<td>70</td>
<td>0.74</td>
<td>-0.95</td>
<td>-0.63</td>
<td>-0.38</td>
</tr>
<tr>
<td>75</td>
<td>0.19</td>
<td>-2.66</td>
<td>-1.36</td>
<td>-0.88</td>
</tr>
<tr>
<td>80</td>
<td>0.19</td>
<td>-7.75</td>
<td>-3.95</td>
<td>-2.54</td>
</tr>
</tbody>
</table>
behavior….care needs to be taken when estimating the distribution of welfare gains with an expected utility framework that has substantive predictions so at odds with observed behavior.”

As noted, models structured to provide closed form solutions are possible only in the face of a host of simplifying assumptions. These include assumptions concerning a complete market—financial products and strategies span all risks; the nature of the equity risk premium—often static and independent of other state variables such as interest rates; and, not surprisingly, the form of the utility function. The majority of models restrict the investment portfolio to two asset classes (risky equity plus a risk-free bond) in which portfolio returns evolve according to a Brownian motion process. The 2003 paper authored by Francesco Menoncin and Olivier Scaillet (“Mortality risk and real optimal asset allocation for pension funds”) provides a good example of building a closed form model to explore the nature of the optimal solution to the problem of utility maximization. In this case, the authors solve the well-known Hamilton-Jacobi-Bellman equation for optimal utility and conclude that investors should, with the exception of a hedging component, generally stay the course with respect to an asset allocation based on the Merton Optimum. However, they acknowledge that their conclusions depend on a model that incorporates strong assumptions. This is especially the case for the CRRA assumption: “…it is well known that the value function usually inherits its functional form from the utility function.”

Decomposition of Annuity Costs

Several studies provide a more direct cost estimate by decomposing the annuity contract into its underlying sales, administrative, investment and actuarial elements. For example, at the June 14-15, 2004 Spring meeting of the Society of Actuaries, Susan J. Sell, a consulting actuary with Milliman, provides the following estimates based on a survey of insurance carriers offering Single Premium Immediate Annuities (“Retirement Income Solutions: Payout Annuities”):

---

51 Empirical evidence indicates that the subgroup of well-to-do retirees who are most likely to seek investment advice exhibit Decreasing Relative Risk Aversion [DRRA]. See, for example, Christopher D. Carroll, “Portfolios of the Rich”, NNER Working Paper, July, 2000.

52 A discussion of the Hamilton-Jacobi-Bellman value equation together with its use in dynamic optimization is found in Chapter 14 of Strategic Financial Planning Over the Lifecycle by Narat Charupat, Huaxiong Huang and Moshe A. Milevsky (Cambridge University Press, 2012).

53 The Merton Optimum is derived by multiplying the reciprocal of an investor’s coefficient of risk aversion by a fraction the numerator of which is the risk premium (expected return on the risky asset portfolio – risk free rate) and the denominator of which is the estimated portfolio variance. In the case of log utility, the portfolio exhibiting the highest value for the Merton Optimum is frequently termed the ‘growth optimal’ portfolio.
• “On average, for the longer payout options, the compensation is about 3 percent to 5 percent. For the shorter payout options, it’s 1.5 percent to 3 percent…. The lowest average was in the wirehouse channel at 3.3 percent, and the highest was in the independent producer channel at 4.3 percent.” [P. 4]

• “Expenses are all over the place….On average this is about 80 basis points on the premium side and $235 per contract. Similarly, on the maintenance-expense side, we typically see a per-policy and basis points of assets. Averages were $55 per contract there and 12 basis points.” [P. 4]

• “For common target surplus assumptions, by far the majority of them hold a percent of statutory reserves that’s about 4.3 percent. Some incorporate a premium component, and that averages 3.8 percent of statutory reserves, plus 4.84 percent of premium. On average, these levels represent about 250 percent of NAIC risk-based capital (RBC).” [P. 5]

• “For pricing targets, we did not ask them what actual profitability they were realizing, but by far again, the majority use statutory internal rate of return (IRR) as the pricing measure. It averages about 12 percent. The second most common measure is GAAP return on equity (ROE), and it has similar ranges and averages as the statutory IRR.” [P. 5]

• “We asked our survey participants to report what their average asset mix was for their immediate annuities, and about 70 percent of the assets were in investment-grade corporate and commercial mortgages.” [P. 5]

• “Similar to expenses, the required interest spreads are all over the place. Because you have so many different product designs, some contracts don’t have any loads; some have policy fees, annual loads and upfront percent of premium loads, so it’s difficult to generalize spreads. They ranged in our survey from 50 basis points to 320 basis points….On average we saw a spread of about 118 basis points for a five-year period certain and 89 basis points for a single life option.” [P. 5]

David Weinsier, presenting at the Spring 2005 meeting of the Society of Actuaries [“Hot Topics in Fixed Annuities”], provides further insights:

• “In terms of profit measures, you have the traditional IRR measure….Profit margin is popular obviously, return on assets, GAAP ROE….Most folks are still shooting for that 12 percent IRR.” [P. 10]

• “What about the impact of interest rates, obviously a key moving part to your fixed annuities? I think we all know that low rates cause spread compression. With a low sustained rate, your higher earning assets end by going over, you have to invest them low, that brings your portfolio yield down, and you are going to realize spread compression. On the other hand, a rapid rise in
rates is no picnic either. If we see all of a sudden a very rapid rise, then you would likely get some surrender mediation occurring.” [P. 12]

The 2011 monograph by Jeffrey Dellinger [“When to Commence Income Annuities”] makes several observations on annuity pricing:

- “…fixed immediate annuities contain a ‘spread’ between the earned rate on underlying assets (e.g. bonds and other fixed-income securities in the insurer’s general account portfolio segment backing fixed immediate annuity obligations) and the credited rate on the liability (e.g. fixed immediate annuity reserve). This spread, in essence parallels its expense charge counterpart on registered products.” [P. 21 of 54 in internet version]
- The load, when it exists, is typically a ‘percentage of premium’ charge used to cover acquisition expenses such as wholesaler compensation, financial advisor sales compensation, policy issuance, record set-up and other policy acquisition expenses as well as state premium tax, if applicable”54 [P. 27 of 54 in internet version]

When compared to the benefits generated by an actuarially fair annuity, a variety of costs decrement the value received by investors buying an annuity contract in the marketplace.

Finally, in 2012, Norma L. Nieelson [“Annuities and Your Nest Egg: Reforms to promote Optimal Annuitzation of Retirement Capital”] observes that most annuities are sold by firms managing large blocks of life insurance. Some of these companies may be able to offer better than expected annuity returns because “The sale of an annuity reduces the overall risk faced by the firm and produces a corresponding reduction in the needed (risk-adjusted) rate of return.” [P. 8]

**Are Annuities Risk Free?**

Annuities are not risk-free assets because they depend on the issuing carrier’s financial ability to fulfill the terms of the contract. Excellent treatments of this issue are found in the 2007 essay by David F. Babbel and Craig B. Merrill [“Rational Decumulation”], in the 2008 essay by Phyllis C. Borzi and Martha Priddy Patterson [“Regulating Markets for Retirement Payouts: Solvency, Supervision, and Credibility”],

54 Dellinger also presents a case for early annuitization under specific economic circumstances. “Sometimes people believe they should wait to purchase an income annuity until a later date because the mortality credits are higher at higher ages.” Although this assertion is correct, Dellinger asserts that “…purchase of an income annuity at a later date negates income attributable to mortality credits the purchaser could have enjoyed had he or she purchased the income annuity at an earlier date.” [P. 8 of 54 in internet version] In other words, the election to annuitize trades the incremental benefits of marginally higher current income against the possibility of receiving a much greater income entitlement at a more advanced age. In Dellinger’s model, the need for greater immediate income is the primary driver in the decision making process. If there is no need for extra current income, it makes sense to postpone annuitization.
in the 2008 essay by Alicia H. Munnell [“The Role of Government in Life-Cycle Saving and Investing”], and in the 2010 legal essay by S. Andrew Pharies [“Primer on Commercial Annuities for Trust and Estates Attorneys”].

In their 2008 research review sponsored by the Society of Actuaries, [“Longevity Risk Quantification and Management: A Review of Relevant Literature”], Thomas Crawford, Richard de Haan, and Chad Runchey note that annuities are not financial instruments with payoffs orthogonal to the capital markets. They are backed by bond portfolios and are guaranteed by corporations operating within the markets. Bond default risk increases the risk that an annuity will fail. They caution:

- “Currently …insurers tend to back their annuity liabilities with a significant amount of corporate debt. These investments have inherent risk of default, which would leave the insurer with less assets than expected to provide for the annuity payments in situations where economic growth was slowed below expected levels, or the economy was in recession.” [P. 25]
- “Adverse mortality experience, whether higher or lower than expected, has implications for reserving and for capital requirements if the ability of the life industry to raise capital becomes impaired based on the market’s perception of the variability of life company debt.” [P. 23]
- “…the lives that purchase annuities can be very different to those who purchase life insurance. Thus, companies that use natural hedging to manage longevity risk are therefore exposed to basis risk.” [P. 37]

These observations have important implications both for financial advisors and for builders of retirement income risk models. Incorporating the risk of carrier insolvency into risk models can significantly reduce the attractiveness of annuities measured either in utility or dollar wealth space. A good example of incorporating the risk of counterparty default is the study by Gaobo Pang [“Good Strategies for Wealth Management and Income Production in Retirement”].

In his 2013 survey of annuity research [Life Annuities: An Optimal Product for Retirement Income], Moshe Milevsky notes:

- “More than 165,000 policyholders had purchased high-yield annuities from Baldwin-United, and the money was frozen for more than three years while regulators and the courts picked up the pieces.” “Another saga that has been ongoing for 20 years…is Executive Life Insurance Company of New York.” [P. 25]
- “NOLHGA [National Organization of Life & Health Insurance Guaranty Associations] has been active recently in the following instances: when Golden State Mutual Life Insurance was shut down by regulators in California in September 2009; when Shenandoah Life entered receivership
in Virginia in February 2009; when Standard Life Insurance Company of Indiana was taken over by Indiana regulators in December 2008; and when London Pacific Life & Annuity Company was liquidated in July 2004.” [P. 26]

- “…the concept of diversification applies not only to stocks and bonds but also to insurance policies, including life annuities.” [P. 27]

Financial advisors should consider the consequences of a failure to diversify the annuity portfolio in light of both potential carrier insolvencies and limitations in state insurance guarantee funds.55

**Interest Rates, Equity Market Performance, and the Annuity Purchase Decision**

Some studies consider the impact of interest rate changes on the timing of an annuity purchase decision. For example, in 2001 [“Annuity Markets and Retirement Security”] James Poterba questions the wisdom of purchasing an annuity in a low interest rate environment. He cautions that annuitizing all wealth at once is a type of annuity market timing. Annuity yields are a function of interest rates and, given the substantial variation in rates over time, it may be unwise to trade all financial wealth for an annuity income stream based on the interest rate prevailing at any single moment. This is, of course, especially the case when interest rates are low because, all else equal, annuity costs decrease as interest rates rise. Mike Orszag, however, provides a counterpoint in 2002 [“Ruin in Retirement: Running out of Money in Drawdown Programs”] by pointing out that although annuities are expensive in a low interest rate environment, the ratio of annuity income to bond income may be attractive because of the mortality credit. The lower the interest rate, the greater the relative impact of the credit.

An investor may wish to maintain exposure to a portfolio of risky assets in order to exploit the opportunity to achieve an increase in wealth sufficient to lock in a future income in excess of that which an annuity offers today. John Ameriks and Liqian Ren stake out this position in their 2008 study [“Generating Guaranteed Income: Understanding Income Annuities”]: “Given ongoing variability in both prices and spending needs, many people might do better by continuing to share in the risks and returns of a well-chosen and diversified set of investments as a resource for their cash flow. Compared with fixed payments, certain assets may be more volatile but also may provide a far better means to hedge spending risks such as inflation over long periods.” [P. 7]

However, by following such a strategy the investor incurs two risks: (1) the investment returns may be poorer than expected; or, (2) the cost of the annuity might outpace the realized returns on the investment portfolio. In either case, instead of being rewarded for taking risk, the investor may discover that he or

---
55 Best’s Key Rating Guide lists guaranty fund provisions for each state. Part Four of this literature survey (“Annuity Basics”) provides additional discussion on this topic.
she must suffer a diminished future standard of living. Moshe Milevsky and Virginia Young, for example, discuss these risks in detail in 2002 [“Optimal Asset Allocation and the Real Option to Delay Annuitization: It’s not Now-or-Never”]. A decision to delay annuitization in the hopes of improving the future budget constraint by ‘out-earning’ the annuity is equivalent to taking a chance on interest rate movements and equity market performance. Milevsky and Young’s observations are echoed by Stephen Abels, an actuary with Mutual of Omaha. He notes, at the Spring 2005 meeting of the Society of Actuaries, [“Payout and Income Annuities”]: “…people are not as inclined to lock in those low interest rates. They don’t want to lock in the payouts that result from calculating them in low interest rates for the rest of their life.” [P. 11]

A broad cross section of commentators caution investors not to rush into annuity solutions during the current period of historically low interest rates. In 2008, Wolfram J. Horneff, Raimond H. Maurer, Olivia S. Mitchell, and Ivica Dus [“Following the rules: Integrating asset allocation and annuitization in retirement portfolios”] voice the opinion: “Even if interest rates are stochastic, retirees will do well to wait until age 80 in the current low interest rate environment.” [P. 397] Their study explores the interrelationships among interest rates, investor risk aversion and the demand to annuitize. As a general rule, the higher the interest rate, the sooner the retiree elects to annuitize. “However, a more risk-loving retiree will also demand a higher short rate than her risk-averse counterpart. The switching frontier itself is concave because the mortality credit increases over time and replaces cost advantages formerly generated by the related short rate. Interestingly, the advice about when to switch to annuities depends on the current level of the short rate, as it relates to the level of mean reversion. The retiree would likely want to wait until age 80 to annuitize if the short rate is below the long-term mean and mean reversion is anticipated….The lower the risk aversion, the higher the short rate must be to induce the retiree to annuitize her assets.” [P. 405] The elegance and complexity of these observations demonstrates the importance of retirement income monitoring and surveillance programs designed to help investors make intelligent asset management decisions. Financial advisors must ask themselves about the wisdom of recommending annuitization in a historically low interest rate environment given the mean-reversionary tendencies of short-term interest rates.

Alessandro Previtero highlights the risk of annuitizing wealth at a specific moment in time. He calculates in his 2011 article [“Stock Market Returns and Annuitization: a Case of Myopic Extrapolation”] that annuitizing after the market drop of 2009 reduced retirement welfare by as much as 10%. The 2012 article by Nathan Zahm and John Ameriks [“Estimating internal rates of return on income annuities”] illustrates how a change in interest rates can dramatically impact payout amounts offered to contract buyers. For example, for a Male Age 65, the annuity payout amount decreased by 8.85% on the nominal
benefit contract and 6.07% on the inflation-adjusted payout contract between April 20, 2011 and October 27, 2011. Mark J. Warshawsky provides a broader historical view of the risk of trying to time annuity purchases in his 2012 essay, “Recent Developments in Life Annuity Markets and Products.” He writes, “…at the end of May 1984, a $100,000 premium bought a monthly payout of $1,134 for a couple. By the end of June 2003, however, as interest rates fell to secular lows, the same $100,000 bought only $503 in fixed monthly lifetime benefits….By December 30, 2008, the fixed monthly lifetime payment on newly issued SPIAs had dropped to only $417 before recovering throughout 2009 to the $500 level.” [P. 40]
PART FOUR: REVIEW OF ACTUARIAL SOLUTIONS TO MITIGATING LONGEVITY RISK

This section of the literature review provides a broad-brush survey of the U.S. annuity marketplace as of the end of 2014. It constitutes a brief introduction to currently available annuity contracts. For the most part, it does not discuss the literature that evaluates annuities as instruments to enhance utility within a Life-Cycle context, or as benchmarks for monitoring and evaluating the health of a retirement income portfolio. Rather, it is intended to acquaint investors and advisors considering actuarial approaches to portfolio sustainability with the types of products offered by the marketplace.

Many investors have a passing acquaintance with two insurance-oriented products: the life annuity which promises to pay a guaranteed income for the remaining life of the beneficiary; and, the reverse annuity mortgage which promises to provide income—perhaps for life—secured by a lien against a personal residence. The life annuity has been around since the Middle Ages; the reverse annuity mortgage program first appeared in the 1980s in the form of a private insurance plan. The reverse annuity mortgage involves strategies to monetize an illiquid asset, and merits a separate exposition. Therefore, this discussion focuses on the following actuarial products:

- The single premium immediate retirement annuity;
- The single premium deferred retirement annuity;
- The variable annuity with lifetime income guarantee rider; and,
- The ruin contingent deferred annuity.

The variety of annuity contracts and policy riders makes it difficult to formulate universally accurate statements about products actually available to consumers. An investor seeking to annuitize some or all of his wealth should engage a qualified advisor to represent their interests in the marketplace.

---

56 The term ‘beneficiary’ can mean a single individual or a beneficiary plus spouse. Where the term encompasses two individuals, the life span measured is the joint life span. For purposes of expositional clarity, we assume that the annuitant, payer, and contract owner are the same.

57 Merchants often traded personal wealth to local monasteries in exchange for the promise of lifetime support and protection for themselves and their families within the confines of the monastic institution. Many monasteries later found themselves in grave financial crisis because they failed to collect sufficient wealth to fund long-term obligations. The reverse annuity mortgage plans of the 1980s were initially privately insured. It was not until the early 1990s that the Federal Housing Authority initiated government involvement.
Annuity Basics

Before looking at product types, it is helpful to review some basics. An annuity is a contract in which an insurance company promises to make a series of periodic payments—usually defined as a sequence lasting for life—in exchange either for a large single premium collected at the beginning of the contract’s term or for a series of smaller premiums collected prior to the start of the annuity’s initial payment date. An annuity in which the entire premium is collected at the beginning of the contract’s term and which begins payments shortly thereafter is called an ‘Immediate Annuity.’ The purchase of an immediate annuity involves an irrevocable sacrifice of capital in an amount equal to the premium paid to the insurance company. The purchaser trades a sum of money for an actuarially equivalent income stream which, in turn, is reduced by the dollar amount of fees, commissions and expenses charged by the insurance carrier. The lower the sales and administrative costs, the greater is the periodic income, all else equal. The insurance company invests the premiums in the expectation of earning a return sufficient to cover the obligations to which it has committed itself and to generate a profit. Unless there is a special premium refund feature or payment guarantee provision within the annuity contract, the payments cease upon the death of the annuitant(s).

The basic annuity structure sometimes engenders misconceptions. One sometimes hears annuities described as “giving your money to an insurance company that invests in the same stocks and bonds in which you could invest; and, because all payments cease upon death, allowing the company to reap a windfall because an early demise means that you forfeit your money to the insurance carrier.” This incorrect description of the actuarial principles underpinning an annuity misses entirely the concepts of ‘risk pooling’ and ‘mortality credits.’ Here is an example: Assume that eleven investors, each age 80, contribute $1,000 at the beginning of the year so that they can collectively purchase an $11,000 Certificate of Deposit maturing in one year and paying a simple rate of interest of 3%. Each investor has the expectation of receiving $1,000 in principal + $30 in interest one year from today. This calculation assumes that each member of the pool receives a pro-rata return from a CD with a total maturity value of $11,330. Now, suppose the eleven investors decide that if any one fails to survive until the CD’s maturity date, the share due to that investor will, instead, be distributed to the remaining living members.58 If one member fails to survive the requisite period, the remaining members divide the $11,330 CD proceeds into ten shares each of which is worth $1,133. By pooling risk, the survivors have reaped a mortality credit of $103—an extra return of 10.3%. The deceased pool members forfeit their shares to the surviving

---

58 This is a form of contract known as a tontine which is illegal because it may encourage some participants to take steps to assure the premature demise of other pool members. An insurance company, however, pools many individuals into contracts where the death of any one member does not impact the payout promised to the remaining annuitants—i.e., there are no “death dividends” to reward the survivors.
members—not to the bank which issued the CD. The issuing bank paid a 3% rate of interest which was sufficient to attract capital that, in turn, was loaned out to other bank customers at a higher rate of interest. The bank’s profit expectations are already built into the CD contract and the bank remains indifferent as to the fate of individual pool members. Likewise, an insurer guarantees the annuity payout for each contract irrespective of which annuitants—or how many annuitants—survive during the period. It can do this because of annuity pricing principles. It sets the payouts so that they appear sufficiently lucrative to attract capital (premiums), and are sufficiently conservative to reserve a slice of the expected mortality credits for insurance company profits.\(^5\) If an investor is indifferent about making future gifts or bequests, the annuity contract will, assuming reasonable fees and expenses, always pay an annuity yield higher than the interest rates on comparable fixed income investments promising a full or partial return of capital to the investor.\(^6\)

One further item is noteworthy—*an annuity is not a risk free investment*. Annuity payout guarantees are only as good as the insurance company that backs them. Historically, as discussed in Part Three, there have been several insurance company insolvencies which have resulted in annuity payment delays or even in a loss of a portion of the investment in the contract when annuity payments ceased. Annuities are guaranteed only up to a limited value specified by state insurance guarantee funds.\(^6\) Investors should also be aware that the financial guarantees are not a direct obligation of the state, but rather are the pro-rata obligations of individual insurance companies who market within the state.

An annuity has unique tax provisions which investors should carefully consider prior to making a purchase decision. Generally, an annuity enables the investor to avoid recognizing income on any accumulations remaining in the contract. Thus it provides a “tax shield” enabling funds to accumulate in

\(^5\) Technically, profits or losses emerge over time as experience dictates whether the liability reserve for the annuity payments was conservative (profits) or inadequate (losses).

\(^6\) It is, however, misleading to compare the annuity yield to the current yield paid by a bond or a CD. The latter instruments usually involve a return of principal. The replicating portfolio for an annuity payout, however, is a series of zero-coupon bonds with a 100% probability of eventual default. This is truly an oranges to apples comparison.

\(^6\) In California, for example, the current (2014) amount of contract ‘non-forfeiture’ value that is covered under state guarantee fund’s provisions is 80% of the present value of annuity benefits including net cash surrender and net cash withdrawal values up to a maximum of $250,000. The rules are complicated—coverage may apply to each annuitant in a joint contract; benefits may be subject to interest rate adjustments, changing residence to a new state may trigger different benefit levels, etc. Although it is wise to diversify the annuity portfolio over several insurance companies unless the annuity is of a type that is segregated from creditor claims against the insurer, the California Guarantee Association limits total coverage for any one individual to $300,000 assuming the annuitant owns contracts issued by multiple companies. [http://www.califega.org/faq.cfm/] Independent rating firms upgraded the insurance industry outlook from ‘negative’ to ‘neutral’ following the recovery from the global recession. However, several carriers with a large share of the U.S. annuity market have recently exited the marketplace because the financial guarantees embedded in annuity contracts threatened the companies’ financial condition. In the 1990s, both Japan and Europe saw multiple carrier insolvencies as mispriced guarantees thrust insurance companies into dire financial straits.
a tax-favored environment. Additionally, each periodic payment is split into a tax-free return of principal part and a reportable income part. The tax characterization of periodic income follows a complicated set of rules enumerated in the U.S. Revenue Code. The portion of income deemed to be taxable is subject to ordinary income rates. Once the aggregate return of principal exceeds the basis in the contract—i.e., the amount of premium paid—all distributions are fully reportable ordinary income. Proponents of accumulation-oriented annuities point to the advantages of the tax shield that allows the investment to grow on a tax-deferred basis. Critics of annuities—both accumulation annuities and immediate payout annuities—point to the tax law provisions that convert investment gains, usually subject to low capital gains rates if held outside an annuity, to investment gains taxed at higher ordinary income rates when received from an annuity.

The Single Premium Immediate Retirement Annuity

The single premium immediate annuity [SPIA] is a contract that begins periodic payments shortly after collection of the initial lump sum premium payment—usually within 30 days. Assuming sufficient financial wealth, the attraction of a single premium immediate annuity is the ability to lock in a lifetime payment stream, perhaps on an inflation-adjusted basis. This type of transactions represents an exchange of risk—not an elimination of risk. That is to say, the annuity purchaser voluntarily assumes counterparty risk—the risk that the insurance company will be unable to honor its contractual payment obligations. The investor can lock in the lifetime income either upon the commencement of retirement—assuming sufficient funds; or, if the investor prefers to wait and see how events unfold in the capital markets, at a later date assuming that sufficient wealth remains to cover the annuity’s future purchase price.

Although there are many ways to use SPIA contracts to generate retirement income, most are variations on two basic approaches—as a product to secure a base level of income; and, as a safety net should investment results prove unsatisfactory:

1. The investor determines the minimum income required to sustain threshold expenses during the investor’s lifespan. This minimum threshold may be less than the “aspirational” amount of income that the investor hopes to have available to spend. Nevertheless, the idea of locking in a threshold income amount at the start of retirement may appeal to certain investors. Provided that the annuity contract is reasonably priced, the annuity’s risk pooling and mortality credits make the attainment of lifetime income cheaper than the income stream available from government-guaranteed bond portfolios. In terms of financial economics, the annuity portfolio’s pricing advantage “crowds out” the bond component of the portfolio.

---

62 The tax regulations for inflation-adjusted annuities differ, in some important respects, from those applicable to nominal payout annuities.
The portion of portfolio assets not used to purchase the annuity income stream may be invested in a risky-asset portfolio with the expectation that the portfolio will generate future capital growth.

2. An alternate SPIA strategy is to consider the annuity in terms of a safety net against a decline in wealth of such magnitude that it jeopardizes the portfolio’s ability to provide a sustainable and adequate lifetime cash flow. Under this approach, the investor delays purchasing an annuity in hopes that traditional stock and bond investments will generate returns at or above those required to fund both threshold and aspirational consumption objectives. However, should the portfolio encounter a bear market of sufficient severity to compromise financial objectives, he can exercise the option to annuitize. The option to annuitize at a later age may make it cheaper to purchase the annuity income assuming that interest rates and insurance company reserving requirements are favorable at the time the option is exercised. 63

Although the concept of guaranteed lifetime income is attractive, many investors never exercise the option to buy a single premium immediate retirement annuity. Given the annuity’s pricing advantages relative to more traditional bond investments, economists refer to the lack of widespread public ownership of annuity products as “The Annuity Puzzle.” Theory dictates that the product should be popular, reality says otherwise. There has been roughly forty-five years of research in an effort to solve the puzzle. Some of the reasons put forth to explain why annuities are not more widely utilized are:

- Distrust of insurance companies and insurance sales representatives;
- Presence of a threshold inflation-adjusted income stream provided through social security;
- Irrevocable loss of capital upon purchasing of the lifetime income stream;
- Importance of gifting and intergenerational bequest planning for retired investors, or importance of a remainder interest in an irrevocable trust context;
- A strong time preference for consumption suggests that current wealth should not be allocated to an annuity to provide long-term funds when there is only a remote contingency that income might be needed at age 99. Rather, the optimal consumption spending budget shows high early expenditures during a period with a high survival probability, and lower expenditures later in life at a planning horizon with a lower survival probability;

63 It is cheaper to buy a lifetime income of $x dollars at age 75 than at age 65, all else equal. Generally speaking, only investors who consider themselves to be in good health find SPIA contracts to be of interest. If the investor does not expect to enjoy a long life span, longevity risk is a less important factor in lifetime income planning. A few carriers write SPIA contracts with high lifetime payouts for individuals in poor health. Actuaries call these contracts “substandard annuities.” When these annuities are customized to compensate individuals who have been awarded court judgments for life impairing injuries, they are known as “structured settlements.”
• Fear of unexpected “liquidity shocks” such as extraordinary medical expenses that would require large reserves of liquid capital;
• Fear of locking in a permanent budget constraint if most wealth is exchanged for annuitized income; and,
• Fear of ceding control of wealth with the attendant loss of the ability for discretionary spending.

Clearly, the SPIA is a powerful financial tool; however, any purchase decision must involve careful planning and informed consideration.

During times of lackluster economic performance interest rates tend to be low. However, these are precisely the times when the insurance industry highlights how yield-starved investors can capture attractive cash flows by buying annuities. Relative to the income currently thrown off by CDs and government guaranteed bonds, an annuitized income stream seems almost too good to be true. However, the intelligent investor also realizes that when money is paid to the insurance carrier, the carrier must invest in the same low-yield capital markets faced by all investors. All else equal, when interest rates in the general economy are low, insurance carriers do not offer the amount of lifetime income per premium dollar that they offer during periods of higher interest rates. This makes sense because if the insurance carrier can earn only 3% on its assets during poor economies, they cannot pay out as much as they could if their invested assets earn 6%. Yet it is during bad economies that annuities look best to those concerned with inadequate periodic income. In terms of timing, buying an annuity contract in a low-interest / recessionary economy is probably the worst time to do so. This is ironic—the product looks best when the time to buy is the worst! 64

The annuity product provides lifetime cash flows and, therefore, exhibits interest rate sensitivity [bond duration risk measure] higher than many fixed income investments. Thus, when considering the factors contributing to the “annuity puzzle,” the fear of locking in a permanent budget constraint assumes great significance in light of dynamic future interest rates. Buying an annuity is equivalent to exchanging the risks traditionally associated with portfolios of financial assets for increased interest rate sensitivity. Beta risk is exchanged for Duration risk.

64 Holding premium deposit amounts, age, sex, and annuity underwriting pricing formulae constant, the changes in interest rates over time generates a “term structure of annuity payments.” For example, spending $100,000 to buy an annuity for healthy 65 year old female in 1995 buys a much higher lifetime income than a comparable annuity purchase for a 65 year old in 2012.
The Single Premium Deferred Retirement Annuity

The Single Premium Deferred Annuity [SPDA] is a relatively new annuity product offered by only a few companies. It is also known as an Advanced Life Deferred Annuity [ALDA]. At the time of this essay, we are unaware of any commission-free SPDA contract. This said, the SPDA may appeal to certain investors. Unlike the SPIA, the SPDA provides a lifetime income that starts ten, fifteen, or twenty years after the payment of the premium—hence, the use of the term ‘deferred.’ Under this type of contract, the annuitant receives income only if he is alive at the future income start date. For example, suppose a 65-year-old male investor elects to purchase a $20,000 income stream with a 2% per year benefit step up. The income start date is age 85. If the investor is still alive at age 85, he will receive a yearly income of approximately $30,000 with a continuing 2% annual step up in benefits for the remainder of his life. However, if he is not alive on the designated starting date, he will receive nothing.

Not surprisingly, these features result in significant cost reductions relative to an SPIA. As a hypothetical example, assume that a $40,000 per year income with a 2% step up in benefits for an annuitant at age 65 costs approximately $800,000 for a SPIA contract. The same income received under a SPDA starting at age 85 costs approximately $100,000—a $700,000 cost reduction. The reason for the cost savings are threefold: (1) only approximately half of the annuitants will be alive to collect the benefits; (2) the insurance company can invest the premium for a 20 year period prior to paying out benefits; and, (3) the number of years of expected benefit payments is less at age 85 than at age 65.

This product can be attractive to investors with portfolios large enough to make it unlikely that they will face an income shortfall within the next fifteen to twenty years. However, if the investor survives beyond this period, continued distributions may result in portfolio depletion. Given the above example, an investor with a $1 million portfolio needs to transfer $800,000 to the insurance company to secure a $40,000 per year income with a 2% per year benefit increase under an SPIA contract. This leaves only $200,000 for investments and emergency fund reserves. By contrast, the investor can purchase the comparable SPDA contract and retain $900,000 in financial assets. In this case, the risk is that the $900,000 is insufficient to provide the needed income during the twenty-year interval prior the annuity’s start date. Of course, there is also the risk that the investor will not live until age 85 and will not collect any benefits. However, the cost of off-loading longevity risk through the SPDA strategy requires a greatly reduced initial premium outlay.

The Variable Annuity with Lifetime Income Guarantee

This annuity product is, by far, the most widely owned type of annuity contract. The variable annuity [VA] first appeared in the U.S. during the 1970s. VAs are “fund linked” annuity contracts which, when
used for retirement income purposes, usually require payment of a single, up-front premium. The insurance company offers the contract owner the right to invest the single premium into a menu of fund offerings where the menu consists of proprietary funds or funds managed by one or more unrelated mutual fund companies. The SEC considers VAs to be securities and requires insurance agents to hold the requisite securities licenses and to provide investors with a prospectus. One advantage of these products is that they are not deemed to be part of the general assets of the insurance carrier and are, therefore, not subject to creditor claims in the event of the insurer’s insolvency. In the current marketplace the investor can find both commission-based and no-commission products—the cost of which can vary significantly. A second significant advantage is that VA contracts are liquid. The investor, subject to applicable contract surrender charges, does not irrevocably forfeit the funds to an insurance carrier. If sudden expenses arise, the contract can be surrendered, partially or entirely, to provide for the investor’s needs.

The structure of these contracts is complex and the actuarial formulae that determine payouts can seem byzantine. Basically, a VA contract is a series of one or more pooled investments comparable to mutual funds. The investor allocates the aggregate investment account across a menu of money market, bond and stock funds. Occasionally, VA contracts offer more exotic choices such as hedge funds, commodities, and so forth. The aggregated account is “wrapped” in an annuity bundle so that it qualifies as an insurance product. This is important because, in its capacity as an actuarial instrument backed by an insurance company, the contract can (1) offer the owner the right to annuitize the “fund linked” value of the investment account to provide for a lifetime income; and, (2) offer riders which provide additional financial guarantees for both minimum payoffs upon the death of the annuitant and for minimum dollar income benefits during the annuitant’s lifetime.

Regulators (SEC and NAIC) approved the marketing and sale of VA products based on assurances given to them by carriers that the financial guarantees offered through annuity policy riders would be “incidental,” and that such guarantees would not be sufficiently widespread to exercise a material impact on insurance company profitability. Insurance firms excel when it comes to pricing actuarial guarantees;

---

65 The funds under management of outside investment companies may differ from the similarly-named mutual funds offered to investors. For example, in some cases, the expense structure of the mutual funds may differ from the funds offered through the VA menu.

66 The cost structure of VA contracts is complex and, in the main, falls outside the boundaries of this discussion. This said, cost matters—a lot; and the prudent investor should thoroughly investigate the myriad of implicit and explicit costs associated with this product.

67 NAIC is the National Association of Insurance Commissioners. Insurance companies are regulated primarily on the state level with the chief regulatory officer occupying the office of state insurance commissioner. The fifty state commissioners, in turn, belong to the NAIC which acts as an advisory body regarding the need to keep or amend current regulatory standards.
pricing financial guarantees, however, is a much different ballgame. The NAIC is currently evaluating a new type of annuity contract—the ruin contingent deferred annuity which is the subject of the next section—and, to the consternation of several major VA carriers, is also taking the opportunity to reassess the risks and reserving requirements of a broad range of financial guarantees offered as VA contract riders.

The basic VA contract provides the option to annuitize the investment account to provide for lifetime income. However, the amount of income is not based solely on the dollar value of the account. Rather it depends on the complex interaction of two actuarial formulae: (1) the number of annuity units within the aggregate account—the value of an annuity unit rises and falls with market performance; and, (2) the payout value from each annuity unit which, in turn, is based on the contract’s assumed interest rate (AIR) per unit. If an investor pays a premium of $1 million which is allocated across the menu of mutual funds, the payout for the remainder of the annuitant’s life depends on the present value of annuity unit liquidation over life (a unit value of $1 means that the investor purchases 1 million annuity units) rather than the present value of a dollar-denominated lifetime income steam. In a nutshell, an actuary swaps annuity units for dollars when calculating the payout benefit. Whereas, in nominal terms, a dollar is always worth a dollar, an annuity unit is worth whatever the market says it is worth. This means that the investor receives a steady lifetime income consisting of annuity units; or, equivalently, an unsteady lifetime income of dollars.

A contract’s AIR sets an assumed rate of return on each annuity unit. If the contract has a low AIR—e.g., 3.5%—that means that portfolio returns greater than 3.5% will push the dollar value of retirement income higher. If, however, the contract has a high AIR—e.g., 5%—that means that the portfolio’s return must be greater than 5% before the annuitant receives a payout increase. Not only can the annuitant fail to receive an increase in benefits for the first 3.5% to 5% of positive market performance, the expenses, fees and commissions within the VA are paid by liquidating annuity units. Assuming a 5% AIR with a 3% total annual fee, the income beneficiary forgoes any benefit increase for the first 8% of aggregate portfolio gain. One reason why it was not misleading for VA carriers to tell regulatory authorities that

---

68 Recent events saw AAA-rated firms like AIG Life and Hartford Life requiring federal bailout funds to survive the liabilities that emerged when their financial guarantees were suddenly in-the-money. Hartford Life, the company with the largest share of VA sales, exited the VA market.

69 It is estimated that approximately 75% of all VA contracts sold in 2005 included riders with supplemental financial guarantees. Prior to the recent great recession, insurance carriers were in a race to offer more and more competitive products. Currently, it appears that they have reversed direction, and are now in a race to sell products with increased rider costs and/or decreased rider benefits.

70 On the plus side, a VA contract with a 5% AIR provides an initial benefit per annuity unit that is higher than a contract with a 3.5% AIR. Stated otherwise, a investor owning a contract with a 5% AIR is less likely to see substantial future growth in the dollar value of lifetime income but, he receives a greater initial income. Many VA
annuity guarantees were “incidental” with respect to their earnings and profits because it is highly unlikely that the average investor would opt for annuitizing a VA contract for retirement income purposes. Clearly, if VA sales were to penetrate the ‘senior market,’ dollar-denominated guarantees would have to be offered to investors—hence, the appearance of lifetime income guarantee riders.

VA income riders became widespread in 2002 with the marketing of a Guaranteed Minimum Withdrawal Benefit [GMWB]. Prior to that time, most guarantee riders assured the contract holder that the beneficiary would receive a minimum amount of guaranteed death benefits within a designated time period—possibly measured by actual life span—irrespective of the actual value of the account.71 Although the GMWB rider comes in many permutations, the basic structure is a guarantee that the contract owner can withdraw, over a specified period of time, a pre-set annual amount of money until the aggregate withdrawal amount equals the initial premium paid into the contract—perhaps reduced for fees, commissions and expenses. The amount eligible to be withdrawn might possibly increase at specific future dates to reflect a higher account value on those dates. However, once the aggregate withdrawal amount reached its upper bound, the contract terms provided for no further benefits under the rider. The guarantee did not assure a lifetime income stream.

Starting in 2005, insurance carriers offered a new guarantee rider to VA policy owners—the Guaranteed Lifetime Withdrawal Benefit rider [GLWB]. Since the time of its introduction, this rider has become popular because it does not place an upper bound on the aggregate amount of income payable to the investor.72 Basically, the GLWB rider guarantees that the contract owner can withdraw a pre-specified fraction of the initial premium—adjusted for expenses, commissions and fees—for life. The allowable fraction of the initial adjusted premium is a function of age. Typically, a 65 year old receives a lifetime withdrawal guarantee of 4% of the premium. Often, there are step-ups in benefits at designated future dates provided that the account value net of previous withdrawals, fees, commissions and expenses attains a dollar value higher than the amount of the initial adjusted premium. By 2005 the insurance industry was able to offer investors a contract with the following advantages:

- Liquidity
- Lifetime income
- Control over investment choices and asset allocation

contract holders opt for a contract with a 5% AIR under the theory that a bird in the hand is worth more than two in the bush. Such distinctions, however, are easily blurred during a sales presentation.

71 The death benefit is adjusted for aggregate withdrawals prior to death.

72 Unfortunately, the names of the various riders have not become standardized across the industry. From time-to-time the term GMWB is used to describe a GLWB. Often, companies have proprietary trademarked designations for these riders which further add to the confusion.
• Tax advantages through deferring reportable gains for funds remaining in the annuity contract
• Tax exclusion of a portion of income received as a periodic annuity payment
• Mitigation of the impact of downside market risk through lifetime income continuation
• Ability to participate in market gains if bull markets pushed account values higher.

Given the list of purported advantages, VA sales skyrocketed. However, in the case of a VA contract supplemented with lifetime income guarantee riders, there are so many complicated moving parts it may be difficult to see the forest for the trees.

A full decomposition of a VA contract is beyond the scope of this review. However, investors should be aware of the following items:

➢ The GLWB guarantee is a contingent guarantee—it is “in the money” only if two events occur concurrently: (1) the income beneficiary remains alive; and, (2) the VA contract value falls to zero. If the two events fail to occur, the guarantee provides no benefits.
➢ An investor is unlikely to receive both the downside risk protection and the ability to participate fully in a bull market. Consider the following example: a newly purchased VA contract has an account value of $1 million and provides a GLWB of 4%. The contract states that one year from now, if the contract’s account value is higher than $1 million, the owner can opt to ratchet up the withdrawal right to 4% of the higher value. The force driving the account value higher—the bull market—is offset by the withdrawals, fees and expenses—forces which drive the account value lower. Bottom line—investors receive a benefit from the downside protection withdrawal guarantee only in the event of a simultaneous occurrence of two future contingencies, and receive the benefit of participating in a bull market only if the market’s performance is extraordinarily positive.73
➢ Sometimes, a marketing pitch might suggest that a VA with ‘step-up’ or ‘ratcheting’ provisions can provide inflation protection because of the potential for participation in market

---

73 These observations are not criticisms of the insurance principles underlying the VA contract. Insurance policy buyers hope to receive no benefits whatsoever for the premium payment made to an automobile insurance company because they do not wish to become involved in an accident. However, if the company is charging a premium of $2,000 per year to insure a vehicle that is worth $1,500, most automobile owners would question the wisdom of continuing this arrangement. Gaobo Pang and Mark J. Warshawsky note that modeling a variable annuity with a guaranteed minimum withdrawal benefit is difficult because “The majority (approximately 70 percent) of the VA+GMWB providers….state in their prospectuses that, upon the automatic step-up or the investor-elected step-up of GIB, the contracts will increase, may increase, or reserve the right to increase the annual rider percentage charges, subject to the contract maximum rates. Changes in market conditions may also trigger such fee hikes.” Pang, Gaobo and Warshawsky, Mark J., “Comparing Strategies for Retirement Wealth Management: Mutual Funds and Annuities,” Journal of Financial Planning Vol. 22 No. 8 (August 2009), pp. 36 – 47.
advances. This claim is dubious given the significant and ongoing drag on account values from costs and from withdrawals.

- The tax consequences of receiving periodic payments under a GLWB rider may differ from those generally applied to payments received under an annuity contract. Payments under the rider may be deemed to be fully taxable ordinary income to the extent that there is any gain in the contract. Under certain conditions, the rider may convert all capital gains to ordinary income and may eviscerate the ability to use capital losses to offset other tax liabilities.\(^74\)

- The more risk averse the investor, the greater the appeal of an income guarantee, all else equal. In a VA contract, however, the investor cannot purchase the rider on a stand-alone basis; rather, it must be bundled with an investment program generating costs that may be higher than those incurred outside of the annuity contract. Thus, the guarantee has both an explicit cost—the amount the insurance company charges for the rider; and an implicit cost—the extra costs of investing through the auspices of the insurance company.

- Are the extra costs worth it? This question should be answered on a case-by-case basis. However, some general observations are possible. A VA annuity contract has costs that are absent from non-annuity investment programs. The extra costs, however, provide two benefits: (1) a tax shield and (2) the right to annuitize according to the actuarial formulae described earlier. For an additional cost, the investor can secure a rider providing a contingent guarantee of a dollar-denominated withdrawal right. It is clear that a rational investor would not purchase an annuity with a 2% non-deductible ongoing fee to obtain the benefit of a tax shield on an investment program generating a 1% pre-tax rate of return. Therefore, in an investment environment which exhibits low interest rates on fixed income instruments, the costs of the tax shield are justified if the investment portfolio has an expected pre-tax return in the neighborhood of 5% or higher. This return may be achievable with high-yield (“junk”) bonds—but the annuity guarantees appeal primarily to risk-averse investors who are not attracted to this asset class. Therefore, to justify the costs and to exploit the purported advantages of the annuity program, the investor in a low interest economy must look to stocks. But the higher the allocation to stocks, the more volatile is the account value. At the end of the day, it may be cold comfort to a retired investor that the withdrawal right is “in the money” while the actual account value is zero. When the account value hits zero, the investor is stopped out from any further participation in financial gains, has lost all liquidity,

\(^74\) As of 2014, a working committee of the NAIC recommends that the GLWB rider be reclassified as a “hybrid annuity.” If the recommendation is adopted, it is uncertain how or if this would affect the taxation of benefits received under the rider.
has permanently locked in a budget constraint, has no inflation protection, and is solely reliant on the continued solvency of the insurer.

- The GLWB rider limits a VA contract’s liquidity. If an annuitant makes withdrawals in excess of those permitted under the terms of the rider, the amount of future income withdrawal rights is subject to modification. For example, if a VA contract holder requires a large withdrawal to pay for unexpected expenses, there is usually a pro-rata adjustment in the amount of guaranteed future income provided under the terms of the rider. Additionally, the amount of excess withdrawal may also be subject to contract surrender charges. Suppose that an 85-year-old investor holding a VA contract with a 4.5% GLWB rider has scrupulously kept annual withdrawals to 4.5% of the original premium amount—as required by the terms of the rider. However, because of poor investment performance, the contract with an initial account value of $500,000 is now worth only $100,000. The investor needs to withdraw an additional $30,000 for unexpected medical expenses. Many GLWB riders make a pro-rata adjustment in the guarantee—in this case a 30,000/100,000 reduction in the guaranteed future payments.

- Perhaps the most questionable element of a VA contract is the implicit assumption regarding the nature of the guarantee itself. Assume that the account value has been fully depleted because of a long and highly virulent bear market. The only way for the investor to believe in the credit worthiness of the guarantee is to believe that the forces resulting in a horrible performance for financial assets have little impact on the solvency prospects of the insurance company. It strains credulity, however, to believe that the financial condition of insurance carriers move independently from general economic trends.

The individual components of a VA contract seem attractive when evaluated in isolation. However, the informed investor considers how the components act in tandem. This is a difficult task because of the complexity of the product and the lack of clear disclosure by some product purveyors. It is certain that a typical VA contract’s costs are a substantial long-term drag on investment performance when compared to no-load mutual funds and exchange traded funds. For investment horizons of 20 years or more, the account values within a VA contract can easily be less than two-thirds of those within a mutual fund portfolio with the same initial value and the same underlying asset allocation. This means that 30% to 40% of wealth may be sacrificed to purchase contingent guarantees. Insurance protection is never free.

**The Ruin Contingent Deferred Annuity**

The GLWB rider requires an insurance company to act as both guarantor and investment intermediary. The Ruin Contingent Deferred Annuity [RCDA], by contrast, does not require the layers of fees
associated with investment programs conducted under insurance company auspices. With a RCDA, the investor retains broader—but not unlimited—control over the portfolio of financial assets while buying only a contingent guarantee of lifetime income continuation. The RCDA strategy allows the investor to approximate a ‘buy-a-guarantee-and-invest-the-difference’ type of portfolio management program. Unlike the irrevocable sacrifice of capital required by SPIAs and SPDAs, the RCDA merely requires the payment of an annual fee. Should the portfolio owner fail to pay the fee, coverage terminates without surrender charges. Presumably, an investor encountering a bear market early into retirement can protect the portfolio by paying the annual fee for the income guarantee. If the sequence of market returns is favorable, the investment surplus makes continuation of the guarantee less necessary; if the sequence of returns is unfavorable, the investor may elect to continue paying until the guarantee is in the money, or until the portfolio sufficiently recovers. The RCDA is akin to the GLBW rider; but, the RCDA is available without having to purchase a high cost VA contract.75

These types of contracts are relatively new, and as of mid-2012, only approximately thirty states have approved the contract for sale. Several firms have withdrawn their product from the marketplace but assert that the product specifications are being reworked and that they intend to reintroduce an updated version. Perhaps the largest current market share belongs to a commission-free RCDA sold through a Transamerica Life subsidiary and based on the claims paying ability of Transamerica Advisors Life Insurance Company.76 The product is marketed under the name of ARIA.77 The ARIA product requires that the investor’s portfolio consist of a broad cross section of designated no-load mutual funds and exchange-traded funds. The portfolio is not managed by the insurance carrier, but remains under the investor’s control at the custodian of choice.78 The decoupling of the guarantee from the underlying portfolio means that the contract does not provide a tax shield.

The amount of lifetime guaranteed income is, in part, a function of the annuitant’s age at the time of the income guarantee election date. The cost of the lifetime income guarantee is, in part, a function of age, current interest rates, and portfolio asset allocation. Portfolio risk is strictly limited by capping the allowable weighting of risky assets to maximum percentages. This means that the cost of guaranteeing a

75 The American Academy of Actuaries report ("An Overview of Contingent Deferred Annuities") states: “A CDA is essentially a stand-alone guaranteed living withdrawal benefit....”
76 The insurer, part of the Aegon Americas group of companies, is rated A+ by the A.M. Best Company and AA- by Standard & Poor’s as of June 1, 2012. Aegon N.V. is an international life insurance, pension and investment group based in The Hague, The Netherlands. The parent company or affiliates, however, may not back the guarantee according to disclosure information: “the guaranteed lifetime payments are backed by the claims-paying ability of Transamerica Advisors Life Insurance Company. They are not backed by any other entity....”
77 ARIA is an acronym for Access to Registered Investment Advisors. An actuarial acronym SALB or ‘Stand Alone Living Benefit” program is also used to describe the product.
78 Fees for the lifetime income guarantee are billed directly to the investor and, unlike a VA contract, are not paid by liquidating investment positions from within the covered account.
lifetime income from a high volatility portfolio may be substantially greater than the cost incurred from a lower volatility portfolio.79

The relationship between portfolio volatility and RCDA cost is an area of concern for regulators. Felix Schirripa, chief actuary with the New Jersey Department of Banking and Insurance heads an NAIC committee tasked with a broad-scope review of VA and RCDA guarantees. The committee’s concern cuts in two directions:

1. What are prudent reserving requirements for such guarantees? This addresses the issue of solvency risk to the insurance carrier.
2. What is the value of the guarantee to the consumer if the insurance company caps portfolio volatility by limiting the proportional weighting of risky assets within the portfolio? This addresses a variety of consumer protection issues.

Schirripa’s actuarial risk models suggest that the annuitant may not receive meaningful or cost-effective longevity protection because of the volatility / cost-of-protection relationship. Schirripa’s work highlights the importance of considering all contract provisions operating in tandem rather than separately. The RCDA annuity product seems to share the same liquidity difficulties as VA contracts—namely that the guaranteed future cash flows can decrease significantly for investors needing to access their accounts for unexpected current expenses. In addition to the cost vs. volatility question, Schirripa’s committee puts forth a variety of issues. These include:

1. Are RCDA contracts financial guarantees or annuity products?80
2. Are the contracts covered by state guarantee funds; and, if so, do they have nonforfeiture provisions?

Tax rules governing CDRA contracts are not yet fully determined. We understand that the IRS has issued a series of private letter rulings suggesting that they will treat income received from CDRAAs as “annuity income” subject to the applicable annuity taxation rules in the Revenue Code.81 However, some

---

79 The ARIA product is offered by prospectus which investors should study carefully in order to understand contract provisions and options. These provisions include the ability to exercise options via “step ups” or “ratcheting” formulae, payout adjustments in the event of large interest rate moves, maximum future fee adjustments, and discounts for large portfolios. Although perhaps not as complex as VA contracts, RCDAs give investors an abundance of fine print to digest.

80 The NAIC Working Committee recommends that GLWBs be classified as ‘hybrid income annuities’ and that CDRAAs be classified as ‘synthetic hybrid income annuities.’

81 Private Letter Rulings apply only to the specific taxpayer to whom they are issued.
commentators have opined that certain withdrawals under CDRA contracts will be taxed at lower capital gains rates. Likewise, the jurisdictional bounds of the SEC remain uncertain.

The probability that the company making the guarantee will remain in business is also of more than passing interest to the investor. The American Academy of Actuaries indicates from a consumer’s perspective, CDAs can be a beneficial product because “CDAs are annuity products which transfer both investment risk and longevity risk to the insurers who issue them.” The CDA Working Group continues to work to establish regulatory recommendations with respect to company financial solvency requirements, regulatory authority, and consumer protections. CDAs are still a work in progress.

It appears as if the RCDA product represents, for some investors, a significant advance over the GLWB/VA package. However, there are a variety of issues remaining to be settled. The informed investor should be fully aware of the NAICs work agenda. Fortunately, it appears that many uncertainties may reach satisfactory resolution in the near future. At that point, the decision whether to acquire the product to protect some or all of the portfolio’s cash flow generating ability will become a more straightforward cost/benefit analysis.

---

82 ARIA marketing material states that “benefit payments are subject to ordinary income tax” if paid under the terms of the ruin contingent deferred annuity guarantee. Additionally, the material states that “the annuity has no cash value, surrender value or death benefit.”

83 The NAIC update of February 26, 2015 notes that a CDA contract “…establishes a life insurer’s obligation to make periodic payments for the annuitant’s lifetime at the time designated investments, which are not held or owned by the insurer, are depleted to a contractually-defined amount due to contractually permitted withdrawals, market performance, fees or other charges.” [http://www.naic.org/cipr_topics/topic_contingent_deferred_annuities.htm]

The CDA Working Group, addressing the issue of whether CDA contracts are fixed or variable annuities promulgated, in the March 24, 2015 Draft “Guidance for the Financial Solvency and Market Conduct Regulation of Insurers who offer Contingent Deferred Annuities [http://www.naic.org/documents/committees_a_contingent_deferred_annuity_wg_exposure_cda_guidance_clean.pdf]: “Because a CDA shares qualities of both a fixed and variable annuity, the Working Group concluded that a CDA should not be classified in either category but instead belongs in its own category.” [P. 7]. A good recap of regulatory concerns with CDAs is found on the February 3, 2013 online edition of the Wall Street Journal: http://online.wsj.com/news/articles/SB100014241278873232468604578247692560852924

84 For example, the investor might compare the current cost of an SPDA where the payout is contingent only upon survival, with the present value of the yearly cost of an RCDA promising to pay a comparable income where the payout is contingent upon both survival and complete portfolio depletion.
CONCLUSION

The literature on longevity risk and portfolio sustainability is, to a great extent, a history of risk modelling. Throughout the fifty years covered in this literature survey, researchers develop a set of decision-making tools to explore, under conditions of uncertainty, dynamic relationships among key variables. Often, there is a twin goal of understanding the nature of these relationships and of suggesting salutary asset management strategies for retirement income portfolios.

One important—and still emerging—theme is that asset management requires credible monitoring and surveillance policies as complex portfolios, operating under conditions of stress, present critical elections for investors seeking to adapt to evolving personal and economic conditions. Modern commentary on this subject tends to view retirement income portfolios as bundles of asset management elections that may be intelligently assessed and implemented either to mitigate risk or capitalize on investment opportunity. Indeed, to a certain extent, a retirement income portfolio can be viewed as a stream of cash flow targets plus a set of management options.

Early published research defined important issues and began a critical examination of important topics. Often, however, early models focused on only a few variables of interest and, given computational restrictions, used simple inputs such as a stylized two-asset investment portfolio generating log-normal return distributions under a constant inflation process. Sometimes, decision making, in the form of spending rules, was set in stone at the beginning of the planning horizon. No matter how events unfolded in the future, the linear, closed-form solutions demanded by a risk model’s underlying mathematics barred any contingent decision-making activity. Future investment decisions are fixed at the outset; circumstances may change but the model’s structure does not. Portfolio monitoring and surveillance, by definition, is of secondary importance in such a highly artificial context. Normative models concentrated on providing asset allocation design and implementation advice capable of withstanding the stress of pre-set spending targets. In some cases, research efforts focused on the elusive task of discovering ‘all-weather’ (a Minimax decision analysis) investment and withdrawal policy. Other studies recommended loading for equity if the investor needs to provide ample money for a lifetime. These approaches are static or architectural approaches to investment policy; and they are often coupled with asset allocations that maintain a fixed and permanent vector of investment weights.

Later research moves, in some cases, away from this paradigm towards a view that sees asset allocation as the economic contribution of investments thoughtfully arranged to capitalize on unfolding events. The legal profession expresses this concept of prudent asset management with the phrase: ‘care, skill and
caution.’ The necessity to build in a portfolio review process elevates the significance of portfolio monitoring—that is to say, creates the precondition for contingent decision making between investment and actuarial solution paths; or among various financial management solution paths.

The nature of future uncertainty is itself uncertain; and, depending on circumstances, will reveal dangers and opportunities unexpected at the moment of retirement. At first blush, it seems that risk-averse investors, valuing an early and ‘complete’ resolution of uncertainty, may find that annuitization is the utility-maximizing option. However, the risk-averse investors may conclude that annuitization is an irreversible decision that both diminishes liquidity and restrains future consumption opportunities. Given such drawbacks, there is a value, *if economically feasible*, to delaying the option to implement the actuarial solution—at least to delay full implementation—until future events resolve the investor’s uncertainty. The risk/reward tradeoff in this context is the reduction of an investor’s exposure to uncertainty v. the irrevocable sacrifice of investment capital and the potential returns thereon.

This brings the discussion to the feasibility topic. Although there are many books and articles on asset allocation strategies and withdrawal formulae for retirement income portfolios, there is only a scant amount of advice on how to monitor wealth to assess whether goals continue to remain attainable. If current portfolio value is less than retirement liabilities, the portfolio is technically insolvent. Investors can hope that things will work out satisfactorily, buy they cannot expect them to do so. Investing encompasses monitoring; and, as private investors are asked to fill the void created by the decrease in guaranteed corporate-sponsored pension income, presenting a clear, unbiased, and credible assessment of the retirement portfolio’s evolving financial health is an increasingly important aspect of managing longevity risk and portfolio sustainability.
BIBLIOGRAPHY


Li, Feng. 2008. Ruin Problem in Retirement under Stochastic Return Rate and Mortality Rate and its Applications Master of Science in the Department of Statistics and Actuarial Science, Simon Fraser University (Spring).


Pang, Gaobo “Good Strategies for Wealth Management and Income Production in Retirement.”


