What Makes a Better Annuity?

Jason S. Scott, John G. Watson, and Wei-Yin Hu

May 2009

PRC WP2009-03
Pension Research Council Working Paper
Pension Research Council
The Wharton School, University of Pennsylvania
3620 Locust Walk, 3000 SH-DH
Philadelphia, PA 19104-6302
Tel: 215.898.7620 Fax: 215.573.3418
Email: prc@wharton.upenn.edu
http://www.pensionresearchcouncil.org

The authors thank William Sharpe, Geert Bekaert, Steve Grenadier, and Jim Shearer for many excellent comments and suggestions. Any remaining errors or omissions are the authors' responsibility. The views expressed herein are those of the authors and not necessarily those of Financial Engines. JEL Classifications: D11, D91, E21, H55, J14, J26. Keywords: Annuities, annuitization, Social Security, pensions, longevity risk, insurance.

What Makes a Better Annuity?

Jason S. Scott, John G. Watson, and Wei-Yin Hu

Abstract

The wide gulf between actual and predicted annuity demand has been well documented. However, a comparable gap exists between the current and ideal annuity market. In a world with costly and limited annuity products, we investigate what types of new annuity products could improve annuity market participation and increase individual welfare. We find that participation gains are most likely for new annuity products that focus on late-life payouts which offer a large price discount relative to their financial market analogues. For example, the marginal utility from the first dollar allocated to a late-life annuity can be several times that of an immediate annuity. Our welfare analysis indicates that an individual's current assets suggest desirable new annuity products since annuities that lower the cost of the existing consumption plan necessarily improve welfare. Finally, we consider the implications for annuity demand if new annuity products ultimately complete the annuity market. Given access to a complete market, we find *all* individuals *only* purchase annuity contracts with a significant time gap between purchase and payout. At a minimum, enough time must pass between purchase and payout to build up a mortality discount sufficient to overcome the cost of creating the contract. Since most existing annuity products, such as immediate annuities, do not have this feature, few current annuity contract configurations are likely to survive significant product innovation. Taken together, our results indicate that there is ample opportunity for innovation to spur annuity demand and improve individual welfare.

Jason S. Scott (Corresponding Author)
Financial Engines, Inc.
1804 Embarcadero Rd.
Palo Alto, CA 94303
T: 650-565-4925 ● F:650-565-4905
E-mail: **jscott@financialengines.com**

John G. Watson
Financial Engines, Inc.
1804 Embarcadero Rd.
Palo Alto, CA 94303
T: 650-565-4922 ● F:650-565-4905
E-mail: **jwatson@financialengines.com**

Wei-Yin Hu
Financial Engines, Inc.
1804 Embarcadero Rd.
Palo Alto, CA 94313
T: 650-565-7771 ● F:650-565-4905
E-mail: whu@financialengines.com

What Makes a Better Annuity?

Introduction

In his seminal work, Yaari (1965) considered an investor faced with two distinct choices—a financial investment and an annuity investment. The singular difference between the options was that the annuity payouts were contingent on survival. Yaari demonstrated that a rational investor should only purchase annuities. This result followed from three fundamental assumptions: (1) an investor's utility depends only on her personal consumption, and in particular, she has no bequest motives, (2) all investments are available with and without a survival contingency, and (3) adding a survival contingency always lowers an investment's cost. Moreover, the estimated welfare gain from annuity market participation was enormous. Full annuitization, assuming access to actuarially fair annuities, often is estimated as comparable to a 50% increase in wealth (Mitchell, et al. 1999; Brown and Warshawsky 2004). In short, economic theory predicts widespread annuity market participation resulting in substantial welfare gains, provided all financial assets are available in cheaper, annuity versions.

Unfortunately, neither Yaari's assumptions nor his predictions reflect the reality of the current annuity market. Consider the assumption that all investments are available with and without a survival contingency. The financial markets offer a staggering variety of investments—dozens of commodities, thousands of different stocks, and tens of thousands of individual bonds. Many of these financial investments support derivative securities that enable individuals to finely tailor their payouts. In contrast, the annuity market is much more limited. Even the most basic securities, such as a guaranteed future payout, are not available individually—but rather must be purchased as part of a standard package of payouts. In addition to limited options, estimates of annuity market costs fall well short of an actuarially fair ideal. While costs for financial assets are often measured in tens of basis points, the cost of adding a survival contingency is typically measured in hundreds, or even thousands, of basis points (Warshawsky 1988; Mitchell, et al. 1999).

Given at least two of Yaari's fundamental assumptions are violated, it is not too surprising that the prediction of full annuitization has not been realized. However, not only do individuals fail to fully annuitize, the vast majority do not participate in the annuity market at all.[1] Unfortunately, violations of Yaari's basic assumptions seem likely to persist. The complexity associated with writing and administering a survival contingent contract requires some level of additional costs and likely implies only a small minority of financial assets will have annuity analogues. However, while Yaari's ideal is unlikely to be reached, that does not mean annuity markets are necessarily destined to languish with low participation rates. Whenever an insurance company lowers the cost of an existing annuity bundle or introduces a new annuity bundle to the marketplace, the annuity market moves closer to Yaari's ideal, resulting in at least the possibility of dramatic increases in participation and substantial welfare gains. However, insurance companies are faced with a nearly infinite array of potential new product offerings. Our research objective is to narrow the field of potential annuity market improvements. In a world of costly and limited annuity markets, we use economic theory to identify those annuity contracts most likely to result in participation and welfare gains. In short, we explore what makes a better annuity.

To analyze participation, we argue one must consider an individual whose wealth is fully allocated to the financial market and examine the marginal utility gained from switching a small amount of wealth to an annuity asset. We demonstrate that this gain is maximized for the annuity product that offers the largest percent discount relative to its financial asset counterpart. Since annuity discounts tend to be proportional to mortality, insurance companies hoping to improve annuity market participation should focus on annuities with late-life, and thus highly discounted, payouts. For questions of welfare, we note that full knowledge of an individual's utility function is required to precisely calculate welfare gains. However, we demonstrate that an individual's investment selection reveals enough about preferences to derive a lower bound on welfare gains. In particular, the welfare gain must be at least as large as the discount the annuity innovation offers on an existing allocation. Hence, insurance companies should look closely at the individual balance sheet when constructing products, since offering survival-contingent analogues to large existing asset holdings should create large welfare gains.

[1] NAVA (2008) reports aggregate immediate fixed annuity sales of $12.8 billion and aggregate immediate variable annuity sales of $0.3 billion in 2005. Given the preponderance of death benefits and premium rebates, it is unclear what fraction of these sales represents actual survival-contingent payouts.

While our goal is to analyze questions of participation and welfare, our first task is to specify a sufficiently robust analytical framework. To match the reality of investment opportunities, we need to introduce both a financial market and an annuity market. While the financial market should be extremely flexible, the annuity market formulation should capture both the potential for significant costs and limited availability. Further, we need a working definition of an improvement to the annuity market. It is from this set of improvements that we will look to find participation and welfare gains. Finally, we will need to identify the (utility) maximization problem we assume characterizes individual behavior. We detail our analytic framework in our first three sections and then turn to questions of participation and welfare in the ensuing two sections.

1. Financial and Annuity Markets

Throughout our analysis, we call all investments without survival contingencies financial assets and refer to them collectively as the financial market. Similarly, the investments with survival contingencies are called annuity assets and belong to the annuity market. For both markets, we assume that an asset's payouts are easily converted to consumption without additional costs, and therefore payouts and consumption are interchangeable. In this section, we describe and model the investment options that we analyze.

A Complete Financial Market

We assume that the current and all future states of our financial market can be described and collected into discrete sets. In particular, we let S_t equal the set of possible states at time $t \geq 0$. The sets $S_0, S_1, \ldots$ are necessarily mutually disjoint. The current state of the world is known, and hence S_0 has only one element. On the other hand, for any future time $t > 0$, there will generally be multiple states with known probabilities that sum to one. Often, we will work with the union $S = S_0 \cup S_1 \cup \ldots$, the set of all financial states. Since every state is a member of just one set S_t, we can define a time function $t(s)$ for every $s \in S$. For simplicity, we will restrict time to discrete, annual intervals. The states describe financial market outcomes, e.g., the cumulative market return equals 80% after five years is an element of S_5. However, no states are allowed to depend on an investor's survival.

Our financial market is complete and defined by a set of Arrow-Debreu contingent claims—one for the current state and one for each of the future financial states. The state-claim on state s $\in$ S pays $1, if and only if, state s occurs. The price of the state-claim, the state-price, is given by the function ψ(s) for each state s $\in$ S. Given access to state claims, an investor can use the financial markets to purchase any tailored package of payouts that she desires—a case of maximal flexibility. A useful package of state-claims is the risk-free, zero-coupon bond that pays $1 at time t, independent of the outcome of the financial market. We assume there are no additional costs associated with purchasing financial market packages, thus the current price Z_t of this zero-coupon bond is simply:

$$(1) \quad Z_t = \sum_{s \in S_t} \psi(s)$$

In particular, the state-claim on the single state at $t = 0$ costs $1, and so we have $Z_0 = 1$. The set of all zero-coupon bond prices describes the financial market's yield curve.

An Incomplete Annuity Market with Frictions

In an ideal world, the annuity market would mirror the flexibility of the financial market. A complete annuity market is exactly the same as a complete financial market with one exception—the investor must be alive to receive a payout. If we assume that the price for the annuity state-claim is actuarially fair, then its price $\gamma(s)$ is given by:

$$(2) \quad \gamma(s) = \pi_{t(s)} \cdot \psi(s)$$

In equation (2), $\psi(s)$ is the price of the corresponding financial state-claim, and π_t is the probability that an investor is alive at time t, given they are alive at $t = 0$.[2] The annuity state-claims are only available for the subset Ω of states for which there is some chance that an investor is alive, i.e., $\Omega = \{s \in S \mid \pi_t > 0\}$. We call these states the *pool set*—the set of financial states an insurer considers when underwriting contingent claims for a

[2] In reality, there are multiple cohorts of individuals, all with different survival probabilities, and all of these cohorts comprise the total annuity market. However, since investors are mostly members of a single cohort, and can only purchase annuities offered to their cohort, we will simplify our presentation by fixing on a single cohort at a time, and one probability of survival. We leave for a future paper some interesting cases that are not covered by this simplification, e.g., a couple may want to explore purchasing individual annuity products versus a joint annuity product.

cohort. For simplicity, we assume that the survival probabilities are strictly decreasing, and satisfy the inequality string: $1 = \pi_0 > \pi_1 > \ldots > \pi_T > \pi_{T+1} = 0$, where $T = max\ \{t\ |\pi_t > 0\}$ is the cohort's maximum possible survival time. In this case, the pool set can be written as the union $\Omega = S_1 \cup \ldots \cup S_T$. Further, since $\pi_t \leq 1$, we see from equation (2) that, with actuarially fair pricing, an investor in the annuity market never pays more for a payout than an investor in the financial market, and more likely, pays less. Hence, investors with no bequest motives should invest only in annuities—Yaari's prediction.

The products in actual annuity markets have limited flexibility—an investor cannot purchase an arbitrary package of annuity state-claims, but only standard packages—the ones offered by the insurance companies. We model an annuity package as simply an arbitrary collection of future payouts. For example, an annuity package might pay an amount $G(s) \geq 0$ in each state $s \in S$. The actuarially fair price for this annuity package γ_G is given by:

$$(3) \quad \gamma_G = \sum_{s \in \Omega} G(s) \cdot \gamma(s)$$

For comparison, we can construct a comparable package of payouts from the financial market. If we ignore packaging and transaction costs, the financial market price ψ_G of a package with payout function G is:

$$(4) \quad \psi_G = \sum_{s \in S} G(s) \cdot \psi(s)$$

In addition to limited availability, we assume annuity packages carry prices higher than their actuarially fair price. We will model additional costs using the *money's worth percentage*. Formally, the money's worth percentage is the ratio of the actuarial fair price of an annuity to its actual cost (Mitchell et al. 1999). Informally, a value of one hundred percent means an investor is getting her full money's worth, and any lesser value models annuity market frictions.[3] For an annuity package with payout function G, its actual

[3] There are a variety of annuity market frictions: (1) asymmetric information on unobserved survival probabilities lead to adverse selection costs in these markets, (2) insurance companies incur administrative and capital costs to create an annuity market, (3) brokerage fees are typically in the range of 3%-10% of the contract value, and (4) some states tax annuity premiums, and for these states, the average tax is about 1.5% of the contract value (Mitchell et al. 1999).

price Γ_G can be written in terms of its money's worth K_G and its actuarially fair price γ_G as follows[4]:

$$(5) \quad \Gamma_G = \gamma_G / K_G$$

Even if annuity markets were complete, survival contingent payouts could still entail additional costs. If we allow for frictions in the pricing of the state-claim annuity described in equation (2), its price $\Gamma_s(s)$ is given by:

$$(6) \quad \Gamma_s(s) = \gamma(s) / K_s = \left(\frac{\pi_{t(s)}}{K_s} \right) \cdot \psi(s)$$

The right hand side of equation (6) illustrates the important tension between mortality discounts and annuity frictions. When annuity market frictions are non-trivial ($K_s < 1$), an annuity state-claim may no longer sell at a discount to its financial counterpart. A discount exists only for the states *s* for which the money's worth exceeds the survival probability ($K_s > \pi_t$). In fact, investors should clearly avoid purchasing annuities in states for which frictions exceed the benefit from pooling mortality risk ($K_s < \pi_t$). Thus, frictions in the annuity market have two important implications. First, full annuitization is no longer necessarily optimal—some states are fundamentally better handled by the financial markets. Second, large frictions imply investors should focus their annuity purchases on states with relatively high mortality.

Sample Fixed Annuity Markets

While our theoretical analysis considers annuity markets with arbitrary packages, we will illustrate many of our results by analyzing four fixed annuity markets: immediate, delayed purchase, longevity and zero-coupon. We refer to these as markets for fixed annuities because the annuity payouts are deterministic. While all of these annuity markets allow for deterministic payouts, the markets are not equal. In fact, as we will detail later, each market in this progression represents an enhancement over the previous annuity markets.

Immediate annuities

[4] Equation (5) puts very few restrictions on the annuity market but does implicitly require that annuity pricing is scale independent—purchasing two units of a given annuity package costs twice as much.

A majority of the research on annuities assumes there is but a single option available—the immediate annuity. For an initial premium, an immediate annuity provides equal payments that begin immediately and last for life. The price of an immediate annuity Γ_I is the ratio of its actuarial fair price to its money's worth parameter K_I. The actuarially fair price of an immediate annuity that pays \$1 annually follows from equation (3) and equation (1) with $G(s) \equiv 1$. Combining these results and simplifying yields:

$$(7a) \quad \Gamma_I = (1/K_I) \cdot \sum_{0 \le t \le T} \pi_t \cdot Z_t$$

As expected, the price of an immediate annuity Γ_I is a function of the annuity market frictions K_I, cohort survival probability π_t, and the financial term-structure Z_t.

Delayed purchase annuities

Many researchers have pointed out that investors typically have an option to delay the purchase of an immediate annuity; i.e., they can set aside funds now to purchase an immediate annuity at a future time.[5] We treat an immediate annuity with this delay option as a separate class of annuities—delayed purchase annuities. Because insurance company policies may restrict sales of immediate annuities to investors below a certain maximum age, there is often a time limit on how long an investor can delay a purchase. Further, for convenience, we assume that all delays are annual. The price $\Gamma_D(\tau)$ of the delayed purchase annuity that begins in τ years and annually pays \$1 is given by the formula:

$$(7b) \quad \Gamma_D(\tau) = (1/K_D) \cdot (1/\pi_\tau) \cdot \sum_{\tau \le t \le T} (\pi_t \cdot Z_t)$$

In equation (7b), the money's worth K_D generally depends on the cohort and the delay τ, but for simplicity we suppress these dependencies. Since the underlying immediate annuity is purchased if, and only if, the investor is alive, the survival probabilities in equation (7b) must be conditioned on survival to time τ, and the factor π_τ appears in the denominator of the price. Strictly speaking, the delayed payout annuity is a hybrid

[5] The following researchers have analyzed the strategy of delaying the purchase of an immediate annuity: Kapur and Orszag (1999), Dushi and Webb (2004), Milevsky and Young (2003), Dus, Maurer and Mitchell (2005) and Horneff, Maurer, Mitchell, and Dus (2006). The last reference has an excellent summary of this literature.

asset—it is an annuity provided an investor lives for τ or more years, and it pays a death benefit equal to its purchase price times the risk-free rate if the investor dies before its conversion to an immediate annuity. Thus its payout function has two components—a survival-contingent payout and a death benefit payout. In particular, the survival payout is $G(s) = 1$ for all $s \in S_\tau \cup \ldots \cup S_T$, and zero otherwise.

Longevity annuities

The longevity annuity is a recent innovation in the annuity product space.[6] MetLife introduced the first of these products in September 2004 (MetLife 2004) and other insurers have followed suit.[7] Simply put, a longevity annuity is an immediate annuity whose first few payments are skipped. In principle, payments could commence at any future date; however, insurance companies often require payments to begin prior to a set maximum age—typically age 85.

The survival payouts of a longevity annuity that begins its payments in τ years are exactly the same as an immediate annuity whose purchase is delayed by τ years, i.e., a delayed purchase annuity. However, the longevity annuity has no death benefit; all of its payouts are survival contingent. The price $\Gamma_L(\tau)$ of the longevity annuity that begins its payments in τ years and annually pays \$1 is given by the formula:

$$(7c) \quad \Gamma_L(\tau) = (1/K_L) \cdot \sum_{\tau \le t \le T} (\pi_t \cdot Z_t)$$

Equation (7c) is very similar to equation (7b), however, there is no need for the π_τ factor, since all survival probabilities are conditioned on survival at $t = 0$ for which $\pi_0 = 1$. In this sense, a longevity annuity price dominates the analogous delayed purchase annuity assuming comparable cost factors.

Zero-coupon annuities

Next, we introduce the zero-coupon annuity, a zero-coupon bond with a survival contingency. The price $\Gamma_Z(\tau)$ of a zero-coupon annuity that pays \$1 in $0 \le \tau \le T$ years is:

---

[6] The industry sometimes refers to this type of contract as "longevity insurance." We have adopted the term "longevity annuity" to be somewhat consistent with this nomenclature.
[7] The Hartford Financial Services Group introduced its longevity annuity in March 2006.

$$(7d) \quad \Gamma_Z(\tau) = \pi_\tau \cdot Z_\tau / K_Z$$

The payout function for this annuity is $G(s) = 1$ for $s \in S_\tau$, and zero otherwise. Though this annuity is not available commercially, it is often a useful benchmark since it allows for any deterministic payout stream to be survival contingent.

To analyze our fixed annuity markets, we will need to make some assumptions regarding the prevailing money's worth percentage. The money's worth percentage for actual annuity products has been estimated to be between 70% and 90% (Mitchell et al. 1999; Warshawsky 1988; Poterba and Warshawsky 2000). While our theoretical results apply to situations where the money's worth percentage is arbitrary and could vary between products, our examples will assume a single money's worth percentage applies across annuity products. Thus, our examples will focus on the impact of all annuity products becoming more or less expensive. Interestingly, the results reported in Mitchell et al. (1999) and Poterba and Warshawsky (2000) are suggestive of a single money's worth percentage for a given cohort.[8] Dushi and Webb (2004) and Purcal and Piggott (2008) argue for a single money's worth percentage and perform their analysis based on this assumption.

2. Annuity Product Enhancements

In this section, we introduce a method for ranking annuity markets. As we have defined it, an annuity market consists of a set of survival-contingent products with known payouts and prices. Though some of these products may have a death benefit, we will focus entirely on their survival-contingent payouts. Further, we only consider annuity markets that are paired with a complete financial market. A financial asset's payments are not survival contingent, but pay either an investor or her heirs. Said another way, a financial asset has both a survival-contingent payout and a death benefit of equal value. If we ignore death benefits, we can think of financial assets as possibly expensive, but very flexible, annuity assets. In fact, investors can purchase any survival-contingent consumption stream they desire, provided they can afford it.

[8] Using 1995 data, annuitant mortality and a corporate yield curve, Mitchell et al. (1999) report money's worth percentages that range from 84% to 86% for men and women between the ages of 55 and 75. With 1998 data, Poterba and Warshawsky (2000) report a range of 82.5% to 86.1%.

We say that an annuity market *A′ dominates* an annuity market *A*, if the following two conditions are met. First, the minimum price of every survival-contingent payout stream purchased using the products in *A′* is never greater than the same stream purchased using the products in *A*. We emphasize that the assets in both annuity markets can be freely packaged with assets from the financial market. Second, there is at least one stream that is cheaper to purchase in *A′* than it is in *A*.

We now formally define annuity market dominance. Let $c(s) \geq 0$ be any consumption stream, and $F(c)$ be the minimum cost of this stream using the products in annuity market *A* and the financial market. If $x(s)$ is the quantity of each state-contingent financial security purchased, and $y(a)$ is the quantity of each annuity product purchased, then $F(c)$ is the optimal objective value of the following cost minimization problem:

$$
\begin{aligned}
&(8a) \quad F(c) = \min \sum_{s \in S} x(s) \cdot \psi(s) + \sum_{a \in A} y(a) \cdot \Gamma_a \\
&(8b) \quad c(s) \leq x(s) + \sum_{a \in A} y(a) \cdot G_a(s), \ s \in \Omega \\
&(8c) \quad x(s) \geq 0, \ s \in \Omega \\
&(8d) \quad y(a) \geq 0, \ a \in A
\end{aligned}
$$

In equations (8), Γ_a is the price and $G_a(s) \geq 0$ is the survival–contingent payout function for the annuity product $a \in A$. Similarly, we let $F'(c)$ be the minimum cost function for annuity market *A′*. We say that the annuity market *A′* dominates *A* if (1) $F'(c) \leq F(c)$ for all streams *c*, and (2) $F'(c) < F(c)$ for at least one stream *c*. Condition (1) ensures that all consumption streams are no more expensive, and condition (2) requires that at least one stream is cheaper.

We seldom compare two distinct annuity markets, rather, we are more interested in the effect new products have on an existing market. If, after adding products to a market, the new market *A′* dominates the old market *A*, we say that the new products *enhance* the old market. Further, since the new market includes all of the old market products, we only need to show that condition (2) holds. If a new set of products replicates the existing set of products, but with higher money's worth, then they enhance the market. However, increasing money's worth is not the only way to enhance a market.

Even an annuity product that sets a new low for money's worth percentage could enhance the market if it covers payouts in states that previously could not be annuitized.

To illustrate market enhancements, we first consider a market of just financial state claims, and introduce immediate annuities. Since the payouts from an immediate annuity are constant, this annuity is an enhancement if, and only if, its price Γ_I is less than the price of a constant payout, risk-free bond ladder; i.e., the money's worth K_I must exceed the bound:

$$(9a) \quad K_I > \sum_{0 \le t \le T} \pi_t \cdot Z_t \Big/ \sum_{0 \le t \le T} Z_t$$

We note that the critical value of the money's worth depends on the cohort's survival probability and the current term structure. Next, we reintroduce immediate annuities, but this time we give investors an option to delay their purchases. If we compare the costs of a delayed purchase annuity and a bond ladder for a constant payout stream starting in τ years, we find that delayed purchase annuity is an enhancement whenever its money's worth K_D satisfies the following inequality:

$$(9b) \quad K_D > \frac{1}{\pi_\tau} \cdot \left(\sum_{\tau \le t \le T} \pi_t \cdot Z_t \Big/ \sum_{\tau \le t \le T} Z_t \right)$$

As long as this inequality is satisfied for at least one value of τ, the option to delay the purchase of an immediate annuity enhances the annuity market. For typical values of the cohort survival probability and the term structure, the graph of the right side of equation (9b) is a U-shaped function of the delay τ. Hence, even though the current purchase of immediate annuity ($\tau = 0$) may not enhance the market, the flexibility of delaying its purchase ($\tau > 0$) may enhance the market.

Suppose we add longevity annuities to an annuity market of delayed purchase annuities. These products will enhance the annuity market if at least one of them is cheaper than both a bond ladder and a delayed purchase annuity. In short, the following condition must hold for at least one value of τ:

$$(9c) \quad K_L > \max\left\{\pi_\tau \cdot K_D,\ \sum_{\tau\le t\le T}\pi_t \cdot Z_t \Big/ \sum_{\tau\le t\le T} Z_t\right\}$$

Finally, we can add zero-coupon annuities to our market. Since these annuities only have financial market analogues, they necessarily enhance the market provided at least one offers a cheaper way to purchase a riskless payout. To be cheaper, the money's worth for a zero-coupon annuity must exceed the survival probability for the given payout horizon.

3. Utility Maximization

We model an investor's preferences with a utility function $U(c)$, whose arguments are the state specific levels of consumption $c(s)$. We assume that our investor has no desire to leave a legacy—this has two consequences. First, the consumption stream $c(s)$ for $U(c)$ only includes the states in the pool set Ω. Second, since all consumption utility is survival-contingent, we can ignore any asset's death-benefit payout. Further, we assume that the partial derivative of U with respect to $c(s)$, U_s, is continuous and positive—more is always better. Also, we require that as $c(s)$ approaches zero, the marginal utility becomes unbounded (see Davidoff et al. 2005). This requirement prevents optimal solutions with zero-consumption states (i.e., starvation states). Finally, we will assume that $U(c)$ is concave for $c(s) > 0$.

In spite of all these restrictions, the class of utility functions that we consider is quite large. It includes functions based on expected utility, e.g., the expected value of a time-separable, power utility function, as well as functions that do not satisfy the axioms of expected utility. Further, our class simplifies the task of maximizing utility subject to linear constraints (e.g., budget constraints). In our case, the first order conditions, the Karush-Kuhn-Tucker conditions, are both necessary and sufficient conditions for the existence of a global utility maximum with positive consumption (see Luenberger 1995).

As an example, suppose an investor with an initial budget of W_0 invests only in financial assets. She can solve the following mathematical program to find her maximum utility U_* and optimal consumption stream c_*:

$$
\begin{aligned}
&(10a) \quad U_* = U(c_*) = \max U(c) \\
&(10b) \quad c(s) = x(s),\ s \in \Omega \\
&(10c) \quad W_0 = \sum_{s \in \Omega} x(s) \cdot \psi(s)
\end{aligned}
$$

The decision variable $x(s)$ is the number of state-claims purchased for state $s \in \Omega$, and since each state-claim pays \$1, $c(s) = x(s)$. Equation (10c) is the budget constraint—the initial endowment W_0 equals the total cost of all state claims. Typically, the total cost only needs to be less than the endowment, however, since the marginal utilities are all positive, the full endowment will be spent. The first order conditions for this problem simplify to:

$$
\begin{aligned}
&(11a) \quad U_s(c_*) = \lambda_* \cdot \psi(s),\ s \in \Omega \\
&(11b) \quad W_0 = \sum_{s \in \Omega} c_*(s) \cdot \psi(s)
\end{aligned}
$$

Here, the Lagrange multiplier for the budget constraint, $\lambda_* > 0$, measures the marginal utility with respect to the initial endowment W_0.

More generally, an investor with access to annuity products will solve the following program:

$$
\begin{aligned}
&(12a) \quad U_* = U(c_*) = \max U(c) \\
&(12b) \quad c(s) = x(s) + \sum_{a \in A} y(a) \cdot G_a(s),\ s \in \Omega \\
&(12c) \quad W_0 = \sum_{s \in \Omega} x(s) \cdot \psi(s) + \sum_{a \in A} y(a) \cdot \Gamma_a \\
&(12d) \quad x(s) \geq 0,\ s \in \Omega \\
&(12e) \quad y(a) \geq 0,\ a \in A
\end{aligned}
$$

In equations (12), the decision variable $x(s)$ is the number of financial state-claims purchased for the state $s \in \Omega$, and the variable $y(a)$ is the number of shares of the annuity product $a \in A$. The survival-contingent consumption $c(s)$ is the sum of the payouts from the state-claims $x(s)$ and all annuities—the contribution from the annuity a is $y(a) \cdot G_a(s)$. Since no future trading is allowed, the current choices for x and y determine all current and future consumption c. Equation (12c) is the budget constraint; the full endowment W_0 is used to purchase assets. Equations (12d) and (12e) are no-short sale constraints for financial assets and annuity products, respectively. Allowing short positions in annuity products is problematic due to the ability to fund arbitrary current consumption by issuing future survival-contingent liabilities. However, short positions in financial assets are often reasonable, e.g., if the present value of all financial assets is always non-negative, then short positions have collateral if the investor dies. Still, we opt for the more restrictive, but simpler, no short sale condition. In the Appendix, we identify and solve the first-order conditions for the mathematical program described by equations (12). The solution prescribes the number of shares of financial assets x_* and annuity assets y_* that an investor should purchase to maximize her utility. Generally, the optimal utility U_* and consumption c_* will depend on which products are available in her annuity market.

4. Annuity Market Participation

The original annuity puzzle centered on Yaari's prediction that a rational investor without a bequest motive would only purchase annuity assets. Since the potential benefits are large, why do so few investors actually adopt the strategy? Some economists argue that actual benefits have been overestimated, while others argue that actual costs have been underestimated. Still, both camps consistently conclude that individuals should invest a significant portion of their wealth in annuities (Brown and Warshawsky 2004; Davidoff et al. 2005). Since Yaari's analysis, the puzzle has evolved—economic theory predicts almost everyone should participate somewhat in the annuity market, but surprisingly, virtually no one does. Davidoff et al. suggested that the reasons are behavioral, and Brown et al. (2008) and Hu and Scott (2007) have investigated this hypothesis. In this section, we focus on what drives annuity market participation and which products would increase it. We examine the predictions for both rational investors and for investors with a behavioral bias against annuity purchases.

Participation for Rational Investors

Fundamentally, annuity assets are purchased because they offer a discount on consumption. For our analysis, we measure this discount as the spending-improvement quotient q identified in Scott [2008]. The q for a particular annuity asset is defined as:

$$(13a) \quad q(a) \equiv \left(\psi_{G(a)} - \Gamma_{G(a)} \right) / \Gamma_{G(a)}$$

The spending-improvement quotient reflects the savings (or additional spending) one could achieve by replicating the payouts of a financial asset with a dollar invested in an annuity asset. We define an annuity market's spending-improvement quotient Q as the maximum q available across all annuity products:

$$(13b) \quad Q \equiv \max_{a \in A} q(a)$$

An annuity market participant is any investor who owns an annuity asset, whereas non-participants own only financial assets. To analyze the participation decision, we must investigate an investor's marginal utility as a function of her wealth α allocated to annuities:

$$(13c) \quad \alpha = \sum_{a \in A} y(a) \cdot \Gamma_a$$

A non-participant's investments are all in the financial market, and so $\alpha = 0$. We can identify participants in the annuity market by first assuming they only have access to the financial market, and then assessing the marginal utility associated with the first dollar annuitized. Participants will find they get positive marginal utility from the first dollar annuitized. The marginal utility for initially swapping money out of financial assets and into annuity assets is given by (see Appendix for details):

$$(13d) \quad dU_* / d\alpha \big|_{\alpha=0} = \lambda_* \cdot Q,$$

where λ_* is the Lagrange multiplier of the budget constraint that appears in equation (10c). Since λ_* is strictly positive from our assumption that more consumption is preferred to less, equation (13d) implies that investors participate in the annuity market whenever Q > 0, or equivalently, whenever at least one annuity product is offered at a discount, $\Gamma_G < \psi_G$. Moreover, we see that marginal utility from participation is proportional to the annuity market's spending-improvement quotient Q. Surprisingly, all investors of a given cohort agree on which annuity product delivers the maximum marginal benefit from participation, even though they will likely disagree on the magnitude of the benefit, since the latter depends on an investor's utility function via the multiplier λ_*.

Given equation (13d), it is not surprising that previous research consistently finds at least partial allocations to the annuity market uniformly optimal. Even though we allow for flexible annuity bundles and a general utility function, we find participation is unanimous provided at least one annuity product is offered at a discount. For our rational investor, the only explanation for a dearth of participation is if the market Q is negative. In other words, if annuitization costs are high enough to overwhelm the mortality discount for all available annuity products then no rational investor should participate.

To assess costs as an explanation for participation, we analyze our fixed annuity markets as viewed by a cohort of women aged 65.[9] For example, consider an immediate annuity with \$1 payouts that begin at age 65. Assuming actuarially fair pricing, the price for this annuity is \$17.96. Purchasing the same payout stream in the financial market costs \$26.00. However, if instead of actuarially fair pricing, this immediate annuity is priced assuming a money's worth value of 69%, then the annuity and the financial price for this payout stream would be equal. A money's worth higher than 69% implies the immediate annuity offers a discount on consumption, and any value below 69% indicates the annuity is inferior to the pricing in the financial market. At this breakeven level of costs, q takes on a value of zero. Since the reported range for money's worth percentages is 70% to 90%, perhaps costs are a key driver explaining the dearth of annuitization. However, equation (13d) requires all annuities have inferior pricing if we are to explain an absence of participation.

[9] All survival probabilities are calculated using the GAM94 mortality table adjusted to 2006. The term structure of interest rates is assumed to be a flat 2%.

Figure 1
Breakeven Money's Worth

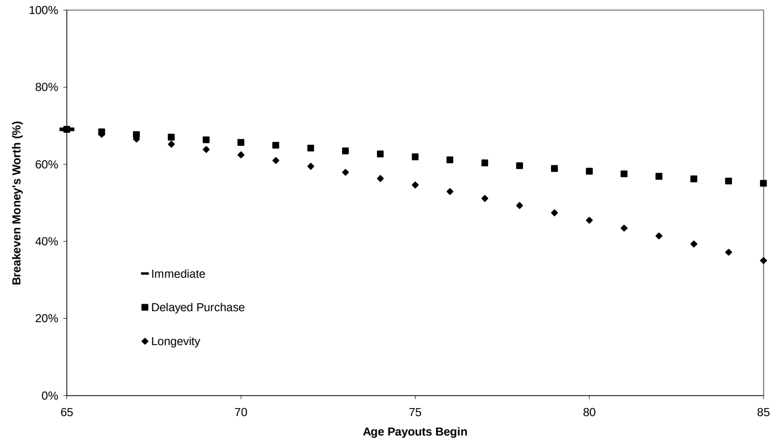


Figure 1 reports the breakeven money's worth percentages for a range of fixed annuity products. The horizontal axis in Figure 1 identifies the age of the first survival contingent payout for the given annuity package. The vertical axis corresponds to the "breakeven" money's worth value that equates the annuity and financial market costs for the given payout bundle. For example, an immediate annuity, a delayed purchase annuity and a longevity annuity are equivalent assuming their respective payouts all begin at age 65. As illustrated in Figure 1, the breakeven money's worth for this annuity bundle is 69%. Data for delayed purchase and longevity annuities are included out to age 85 since this corresponds to the latest age a typical policy allows one to start payments.

Breakeven values for delayed purchase annuities decline from 69% at age 65 to 55% at age 85. Thus, in a high cost environment, it could be optimal to delay the purchase of an immediate annuity. For costs to preclude the use of delayed purchase annuities, money's worth percentages would have to drop below 55%, a value lower than any reported in the literature. Longevity annuities also exhibit declining breakeven money's worth values as the payout age increases. Since longevity annuities price dominate delayed purchase annuities, their breakeven money's worth percentages are lower. By age 85, the breakeven money's worth for a longevity annuity is 35%. In other words, with a longevity annuity, insurance companies must charge a price roughly triple the actuarially fair price to eliminate the discount.

Given our model, a cohort of rational investors uniformly decides whether or not to participate based on the sign of Q. A negative Q is possible given sufficiently high fees. If writing annuity contracts involve high costs which necessitate high fees, then insurance companies should introduce annuities with payouts concentrated in low survival states. This class of annuity packages would allow both a high insurance fee while still providing a discount relative to the financial market. By separating the purchase date from the payout date for all payouts, longevity annuities significantly lower the survival probability for each payout. The result is a contract that offers a discount even in extremely high cost environments. If costs were the main barrier to demand, the introduction of longevity annuities should significantly increase participation. However, since delayed purchase annuities also have a relatively low breakeven and have been readily available for years, costs alone seem insufficient to explain the dearth in participation.

Participation for Investors with a Behavioral Bias

If costs considerations are insufficient to preclude annuitization, perhaps some other factor is involved. In this section, we consider a simple model for investors with a behavioral bias against purchasing annuities. We assume that their utility functions have one additional argument—the amount of wealth invested in annuities α given by equation (13c). In particular, we let $V(c, \alpha)$ represent the utility function of our behavioral investor. Further, we assume that any increase in α decreases utility, i.e., $\partial V/\partial \alpha < 0$. To identify an annuity market participant, we first constrain an individual to only use the financial markets, and then evaluate the marginal utility of the first dollar annuitized. Annuity market participants will find their initial allocation to the annuity market generates positive marginal utility. As we demonstrate in the Appendix, the marginal utility from initial participation is given by:

$$(14a) \quad dV_*/d\alpha\big|_{\alpha=0} = \lambda_* \cdot (Q - P)$$

$$(14b) \quad P \equiv -\frac{\partial V(c_*, 0)}{\partial \alpha} / \lambda_*$$

The annuity penalty P is dimensionless, positive, and depends on an investor's utility function, i.e., it varies within a cohort. We see from equation (14a) that an individual will increase her utility and purchase annuities provided the market's Q exceeds her value of P. While no member of a behavioral cohort will participate unless Q exceeds the cohort's minimum P, all members will participate if Q exceeds its maximum P. However, it is more likely that some members will participate, and the remaining will not. In short, participation now depends on the magnitude of Q rather than just its sign.

Table 1
Annuity Market Q

Money's Worth (K)	Immediate	Delayed Purchase	Longevity
100%	0.45	0.82	1.86
90%	0.30	0.63	1.57
80%	0.16	0.45	1.29
70%	0.01	0.27	1.00

Notes: Interest rates are 2%. Survival probabilities based on age 65 female using GAM 94 mortality tables with adjustments to 2006. The maximum age is assumed to be 100. Delayed purchase and longevity annuities have a maximum payout start age of 85.

Table 1 reports the magnitude of Q for various combinations of fixed annuity markets and money's worth percentages. For example, an actuarially fair immediate annuity has a q-value of $q = (\$26.00 - \$17.96) / \$17.96 = 0.45$. If this is the only annuity available, then this is also the market Q as reported in Table 1. Now, suppose our cohort can also purchase an actuarially fair longevity annuity with \$1 payments that begin at age 85 at a price of \$3.26. The cost of this annuity's replicating bond ladder is \$9.32, and so $q = (\$9.32 - \$3.26) / \$3.26 = 1.86$. If age 85 is the maximum age payouts can begin, then this will also be the annuity market Q.

Assuming a money's worth value of 80%, then historically light participation in the delayed purchase annuity market suggests that most investors have a value of P that exceeds this market's Q, a value of 0.45 (see Table 1). In other words, the behavioral disutility of annuitizing the first dollar exceeds 0.45 times the utility gained from a dollar increase in initial wealth endowment. For the same money's worth of 80%, a smaller behavioral distortion can explain the unpopularity of immediate annuities purchased at age 65: the disutility of annuitizing only needs to exceed 0.16 times the marginal utility of wealth. While introducing new annuities does not affect an individual's P, the marginal utility of participation can increase if the introduction increases the market Q. The introduction of longevity annuities nearly tripled the market's Q to 1.29. If this simple behavioral model is correct, then we should see a boost in participation. The disutility of annuitizing would have to *exceed* the marginal utility of wealth to suppress participation in this market. Moreover, underwriters could increase participation further by introducing new products that increase the market's Q.

While Yaari predicted individuals would spend every last dollar on annuities, participation merely requires some annuity allocation. Surprisingly, we find that the marginal utility from participation is governed *for everyone* by the same quantity, the market spending improvement quotient Q. This suggests that insurance companies looking to improve participation with new annuity products should focus on increasing Q. These new products are not necessarily low-cost, high money's worth products. Indeed, a high Q is most easily attained by offering annuity products with payout bundles focused on high mortality states. Even a relatively high cost longevity annuity market, with a money's worth value of 70%, offers double the Q compared to an actuarially fair immediate annuity.

5. Individual Welfare

Whereas our analysis of the participation decision focused on the marginal utility from an initial allocation to annuities, in this section we investigate the total welfare an investor gains from access to an annuity market and which market enhancements potentially provide the largest gains. We first introduce an investor's expenditure function $E(U_0)$, the minimum cost to achieve the utility level U_0 using the annuity and financial markets:

$$(15a)\quad E(U_0) = \min \mathrm{W}$$

$$(15b)\quad W = \sum_{s\in\Omega} x(s)\cdot\psi(s) + \sum_{a\in A} y(a)\cdot\Gamma_a$$

$$(15c)\quad c(s) = x(s) + \sum_{a\in A} y(a)\cdot G_a(s)$$

$$(15d)\quad U(c) \geq U_0$$

$$(15e)\quad x(s) \geq 0,\ s\in\Omega$$

$$(15f)\quad y(a) \geq 0,\ a\in A$$

For example, consider an investor with an initial endowment W_0 and investments in the annuity market A. Her maximum utility U_* is the solution to the program given by equations (12), and the value of her expenditure function for this level is $E(U_*) = W_0$. Suppose an investor chooses products from an annuity market A' that enhances the market A. We expect that she may pay less, but never more, and still achieve the same level of utility. The money she saves is her welfare gain. Formally, if we let E and E' be the expenditure functions for the annuity markets A and A', respectively, then the welfare gain ΔW is defined as:

$$(16) \quad \Delta W \equiv E(U_*) - E'(U_*) = W_0 - E'(U_*)$$

Informally, an investor's welfare gain is the money she can save by using the products in the enhanced annuity market and still maintain her original level of utility. Given an investor's utility function, we can easily calculate welfare gains.

Portfolio Decisions and Welfare Bounds

Unfortunately, an individual's utility function is not directly observable. Utility functions are individual specific, and while our examples can illustrate effects, welfare calculations based on a particular utility function have limited practical application. In the following, we get around this difficulty by recasting the problem in terms of an investor's observed actions (she makes investments to support her future spending) and away from her unobserved decision process (she maximizes her personal utility function). Our method produces some very general results about how annuity enhancements influence welfare.

Our first step is to rewrite welfare as a function of consumption. If c_* is the optimal consumption computed from minimizing expenses in annuity market A, then $E(U_*) = F(c_*)$, where F is the consumption based cost function defined by equations (8). Similarly, for the annuity market enhancement A', we have the optimal consumption c_*' and expenditure $E'(U_*) = F'(c_*')$. Hence, using these equalities and equation (16), we can rewrite welfare in terms of the optimal consumptions as follows:

$$(17) \quad \Delta W = F(c_*) - F'(c_*') = \left[F(c_*) - F'(c_*)\right] + \left[F'(c_*) - F'(c_*')\right]$$

The second equality in equation (17) splits the welfare gain into two components. The first term in brackets is the money saved by purchasing the original consumption in the enhanced market. We call this the welfare from savings, and since we assume that A' enhances A, it is always non-negative. The second term in brackets is the welfare from shifting consumption streams. Both c_* and c_*' have the same utility U_*, however, in the enhanced market A', the latter is optimal and will never be more expensive than the former.[10] The second term is called the welfare from consumption shifting, and it is also non-negative. Since both terms are non-negative, there is never a welfare loss if an annuity market A' enhances a market A.

Equation (17) relates welfare to consumption—it depends on both the current market consumption c_* and the hypothetical, enhanced market consumption c_*'. However, at the cost of some precision, we can eliminate the latter dependence. Since, the welfare from shifting is non-negative, a lower bound on total welfare is just the welfare from savings, which depends only on the current consumption c_*. Hence, we have:

$$(18) \quad \Delta W \geq F(c_*) - F'(c_*)$$

The sharpness of this welfare bound depends on the size of the welfare from shifting. In general, welfare from shifting depends on the degree to which prices for consumption change and the substitutability of the arguments of the utility function. Since consumption across states is generally modeled with a low degree of substitutability—gorging one year does not make up for starvation the next—the welfare from shifting tends to be small resulting in a sharp bound on the welfare gain. We can make a stronger argument for the quality of this bound in the limit as price changes become small. For small price changes, welfare from shifting vanishes and the total welfare gained is approximately the welfare from savings:

$$(19) \quad \Delta W \approx F(c_*) - F'(c_*)$$

[10] We can easily show that the consumption shift welfare is non-negative. First, the consumption streams c'_* and c_* have the same utility U_*. The cost of c'_* is $F'(c'_*) = E'(U_*)$, but this is the least cost way of getting the utility level U_*. Thus, any other consumption stream with the same utility is at least as expensive, and in particular, we have $F'(c_*) \geq F'(c'_*)$, or $[F'(c_*) - F'(c'_*)] \geq 0$.

In this context, a price change refers to the price differential between an annuity product in the enhanced market and the minimum cost replicating package of financial and annuity products in the original annuity market.[11]

Generally, the bound described by equation (18) requires full knowledge of the optimal consumption bundle c_*; however, we can obtain a weaker bound using partial consumption information. First, consider the investment shares x_* and y_* that are purchased to form the least cost portfolio for c_*. These shares can be arbitrarily parceled out to multiple, virtual accounts. We call these virtual accounts *partial portfolios*, and we call their payouts *partial consumption* streams. If we create M partial portfolios, there will be M partial consumptions $\xi_1, \xi_2, \ldots, \xi_M$, and their sum will equal or exceed c_*. Partial consumption streams have the property that the sum of the welfare savings from each stream is a bound on the welfare savings of their sum, the total consumption.[12] Using this property, we get the following weaker bound on welfare:

---

[11] Consider a simplified situation where two annuity markets, A and A', offer the same products. Annuity market A' is considered an enhancement because it has lower prices. Let p and p' be the price vectors in A and A', respectively. The welfare gain for a fixed utility level U_0 is $\Delta W = e(p,U_0) - e(p',U_0)$, where e is the expenditure function. For small price changes, we can approximate the welfare gain with its differential: $\Delta W \approx dW = \Sigma\, (p_a - p'_a)\cdot \partial e(p,U_0)/\partial p_a$. However, from the Envelope Theorem, we have $y_a = \partial e(p,U_0)/\partial p_a$, where y_a is the number of product shares purchased in market A for utility level U_0. Hence, $\Delta W \approx \Sigma\, (p_a \cdot y_a) - \Sigma\, (p'_a \cdot y_a) = F(c) - F'(c)$, where c is the consumption generated by the portfolio, and F is the cost function.
Surprisingly, we can extend this result to any annuity market A and its enhancement A'. To see this, recognize that we can repackage linear combinations of existing products in A to create new products without altering the expenditure function. In particular, in A we can replicate, at minimum cost, the payouts of any product in A′. We let market B correspond to the union of A and all replicated products of A'. Note that B simply offers some redundant packages, but does not offer any economic advantage relative to A. We can similarly construct B' as the union of A' and least cost linear combinations of products in A' that replicate the products in A. Since economic opportunities are unchanged, welfare gained from B to B' is equivalent to welfare gained from A to A'. However, B and B' offer the same set of products only at different prices, demonstrating that the approximation holds in general.

[12] For annuity market A, suppose we split a minimum cost portfolio for the consumption stream c into two partial portfolios with partial consumption streams ξ_1 and ξ_2. We can show from equations (12) and the definition of partial portfolios that $F(\xi_1+\xi_2) = F(\xi_1) + F(\xi_2)$. In fact, the purpose of partial portfolios is to obtain this equality. Generally, the relation is an inequality, and this is also true for partial consumptions when the costs are computed with a different annuity market, say A′, i.e., $F'(\xi_1+\xi_2) \le F'(\xi_1) + F'(\xi_2)$. From these two properties we get the bound:

$$F(c_*) - F'(c_*) = F(\xi_1+\xi_2) - F'(\xi_1+\xi_2) \le [F(\xi_1) + F(\xi_2)] - [F'(\xi_1) + F'(\xi_2)] = [F(\xi_1) - F'(\xi_1)] + [F(\xi_2) - F'(\xi_2)].$$

This result generalizes for any finite number of partial portfolios and consumptions.

$$(20) \quad \Delta W \geq F(c_*) - F'(c_*) \geq \sum_{1 \leq m \leq M} \left[F(\xi_m) - F'(\xi_m) \right]$$

Each partial portfolio's contribution to the above bound is its welfare from savings and is non-negative. Hence, if any one of the partial consumption streams is unknown, its term can be bounded by zero and dropped from the sum. Given equation (20), we no longer need full knowledge of an individual's portfolio to calculate a bound of welfare gained. Instead, if we can identify any part of an individual's investment portfolio, we can bound the welfare based on the savings the enhanced annuity market offers on replicating that partial portfolio. Of course, as our knowledge of the partial portfolio becomes more and more complete, the welfare bound from equation (20) approaches the bound from equation (18).

The welfare bounds identified in equation (18) and equation (20) have important implications for annuity product introductions. Since welfare gains are primarily savings gains, annuity providers can use a cohort's representative portfolios to design new products. Rather than designing products based on individual utility, underwriters can focus on products that match the payouts of currently held investments. For example, a house represents the single largest asset owned by many individuals. As such, the preceding analysis suggests the nascent market for reverse mortgages could provide large welfare gains for many individuals. Whether the asset is a stock, bond or piece of real estate, the fundamental insight is individual portfolio decisions provide important clues as to desirable annuity product innovations. We next illustrate these ideas with two examples. In the first example, we argue that investments providing constant risk free consumption should receive special attention since this likely makes up a sizeable fraction of many individual portfolios. In the second example, we analyze a representative utility specification to illustrate the quality of the welfare bounds.

Example #1: Welfare Bounds and Minimum Consumption

Equation (20) suggests we focus our attention on any portfolio that is held, either in part or in full, by a large number of people since reducing the cost of this portfolio yields widespread welfare gains. Our utility framework identifies one such portfolio. In our framework, starvation states are explicitly ruled out. Hence, every optimal solution will have a minimum positive consumption value (i.e., a consumption floor). While our framework requires at least some wealth be allocated to minimum consumption, the fraction of wealth allocated could be arbitrarily small. However, a review of the literature suggests wealth allocated to minimum consumption could actually be quite large. For example, in traditional, time-separable utility models, high levels of minimum consumption often result from intertemporal risk aversion or are explicitly imposed with a large subsistence requirement. More recently, habit formation models have developed to explain the surprisingly high levels of consumption persistence. Constantinides (1990) and Sundaresan (1989) pioneered this approach. In addition to the academic literature, consumption smoothing is also a ubiquitous theme in the financial planning literature. Arguably, the most popular rule of thumb for retirement consumption is Bengen's (1994) *4% rule*. This rule pegs retirement spending to a real, constant value—4% of initial wealth. Bengen states that his clients "wanted to spend as much as possible each year from their retirement accounts, while maintaining a consistent lifestyle throughout retirement."[13] In short, there is substantial evidence that many investors want a sizeable floor—a constant, risk-free, consumption stream. Hence, an annuity market enhancement which specifically lowers the cost of floor consumption will likely improve welfare across a broad population.

We can estimate the welfare gain derived from using annuities to support floor consumption using an investor from our sample cohort. Initially, suppose that she invests \$100 in a zero-coupon bond ladder to provide her consumption floor. Table 2 reports her savings for various annuity market enhancements (the columns) and money's worth values (the rows). If all annuities are actuarially fair ($K = 100\%$), then the savings is the same across all markets—the floor purchased for \$100 can now be purchased for \$69.10, a savings of \$30.90. From equation (20), we see that our investor's total welfare gain is at least \$30.90.[14] As the money's worth decreases, early annuity payouts are no longer cheaper than early zero-coupon bond payouts. A longevity annuity selectively avoids

[13] See Scott, Sharpe and Watson (2008) for more detail on the financial planning literature and the 4% rule.

[14] In this literature, welfare gains are often reported in terms of Annuity Equivalent Wealth (AEW). AEW assumes an optimal utility is achieved with annuities and then calculates the additional wealth required to compensate for the loss of annuity access. This approach relies on optimizations in the enhanced market which does not allow a welfare calculation based on observable portfolios. We avoid this problem by returning to the fundamental definition of welfare – equation (16). There is generally not a translation from the reported welfare gain and AEW. However, if the utility function is CRRA and separable, then AEW = \$100 / (\$100 – ΔW). Thus for a \$30 welfare gain, AEW is \$100/\$70 ≈ \$143.

these expensive annuity payouts, and in combination with bonds, it saves more money on a floor than an immediate annuity. Delayed purchase annuities are superior to immediate annuities, but inferior to longevity annuities. Further, the welfare gains of longevity annuities are robust to annuity market frictions, e.g., at the 70% money's worth level, where the welfare gain from an immediate annuity is negligible, the gain from a longevity annuity still exceeds half of its actuarially fair level. Enhancing the market further by providing zero-coupon annuities does not improve upon longevity annuities at these cost levels. Money's worth levels would have to dip below the age 85 survival probability (~63%) for zero-coupon annuities to provide a cheaper path to minimum spending.

Table 2

Welfare Bound per $100 Allocation to Minimum Consumption

	Annuity Market			
Money's Worth (K)	**Immediate**	**Delayed Purchase**	**Longevity**	**Zero-coupon**
100%	$30.9	$30.9	$30.9	$30.9
90%	$23.3	$23.3	$25.4	$25.4
80%	$13.7	$14.7	$21.4	$21.4
70%	$1.3	$8.4	$18.0	$18.0

Notes : Access to the finanical market alone is the base case in all situations. Interest rates are 2%. Survival probabilities based on age 65 female using GAM 94 mortality tables with adjustments to 2006. The maximum age is assumed to be 100. Delayed purchase and longevity annuities have a maximum payout start age of 85 while zero-coupon annuity payouts are available up to age 100.

Example #2: Welfare Bounds and Habit Formation Utility

We have presented two bounds on the welfare gain—the savings welfare bound and the floor welfare bound. The floor bound depends only on an investor's floor consumption; however, if we know all her portfolio's payouts, we can calculate the welfare from savings and improve that bound. Finally, if we know an investor's utility function, we can calculate her actual welfare gain. The bounds are often good estimates of the actual welfare, and in this section, we illustrate their accuracy with a numerical experiment.

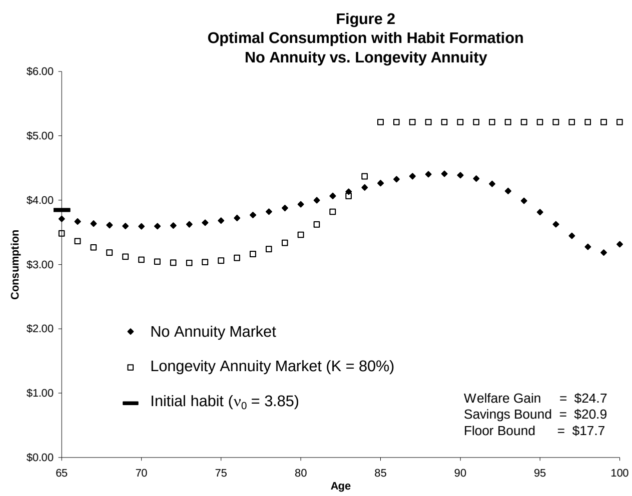
Consider an investor from our sample cohort with the following utility U:

$$\text{(21a)} \quad U(c) = \sum_{t=0}^{T} \pi_t \cdot (1+\delta)^{-t} (c_t / v_t)^{1-\varphi} / (1-\varphi)$$

$$\text{(21b)} \quad v_{t+1} = (v_t + \beta \cdot c_t)/(1+\beta)$$

This habit formation utility was introduced by Diamond and Mirrlees (2000) and later analyzed by Davidoff et al. (2005). Note that consumption c_t at time t is independent of state, i.e., risk-free. In equation (21a), π_t is the cohort's survival probability, δ is a time-discount parameter, φ is a risk-aversion parameter, and ν_t is the habit level. The habit level changes with time following equation (21b) and depends on its previous level, the previous consumption, and the habit persistence β. For our numerical examples, we use $\varphi = 2$ and $\beta = 1$ (Davidoff's values) and a real interest rate of $\delta = 2\%$. We set the initial habit level ν_0 to \$3.85, the floor level our cohort can purchase with a bond ladder.[15]

In Figure 2, we plot two optimal consumption streams versus age for an investor with the above habit utility and parameters. The solid bar at age 65 corresponds to the initial habit level $\nu_0 = \$3.85$. The first stream c_* (solid diamonds) is optimal for an initial endowment of \$100 and when our investor has no access to annuities. The second stream c'_* (empty squares) is optimal for an initial endowment of \$75.3 and when our investor has access to a longevity annuity market with $K = 80\%$. The endowment for c'_* was chosen so that the optimal utility for both streams is the same, i.e., $U(c_*) = U(c'_*)$. Hence the endowment difference, \$24.7, is the welfare gain from introducing the longevity annuity market.



[15] We use a bond-funded floor to fix the initial habit level, whereas Davidoff et al. (2005) used levels of 0.5, 1.0 and 2.0 times an immediate annuity funded floor. We also investigated a range of values, and the results of those experiments are very similar to those reported here.

We also see that for this situation, the two welfare bounds provide a good estimate of total welfare. The introduction of longevity annuities significantly changes the price of consumption across the various ages. The result is fairly substantial shifts in consumption. Consumption prior to age 84 declines on average by 12.5%. Consumption at age 84 and beyond increases by an average of 31%. Even with relatively large shifts in consumption, the welfare from savings is still a good bound for the total welfare gain. With longevity annuities, the original consumption stream can be replicated at a cost savings of \$20.9, thus the welfare from savings captures about 85% of the total welfare gain. Prior to the introduction of longevity annuities, the minimum consumption level was \$3.19. The cost savings on this floor consumption was \$17.7, capturing approximately 72% of the total welfare gain.

Table 3 reports the welfare gain for four annuity markets (column groups) and various values of the money's worth (rows). Welfare triplets are reported for each combination—actual welfare, savings welfare, and floor welfare. For all welfare calculations, an annuity market is considered as an enhancement over the financial market. For example, the actual, savings, and floor welfare gains for an actuarially fair (*K = 100%*) immediate annuity are \$30.4, \$28.4, and \$25.6, respectively. Similarly, the triplet for the longevity annuity (*K = 80%*) is \$24.7, \$20.9, and \$17.7—these values correspond to the consumption streams discussed above and plotted in Figure 2.

Table 3
Welfare and Welfare Bounds with Habit Formation Utility

	Annuity Market											
	Immediate			Delayed Purchase			Longevity			Zero-coupon		
Money's Worth (K)	U	Savings	Floor	U	Savings	Floor	U	Savings	Floor	U	Savings	Floor
100%	**30.4**	28.4	25.6	**31.7**	29.7	25.6	**34.7**	30.8	25.6	**36.9**	31.3	25.6
90%	**22.6**	21.2	19.3	**24.1**	22.0	19.3	**29.2**	25.2	21.1	**31.7**	25.8	21.1
80%	**13.1**	12.2	11.3	**15.6**	13.5	12.2	**24.7**	20.9	17.7	**27.3**	21.5	17.7
70%	**1.2**	1.1	1.1	**8.8**	7.2	6.9	**20.6**	17.2	14.9	**23.2**	17.9	14.9

Notes : Access to the financial market alone is the base case in all situations. For each annuity regime / money's worth combination the welfare gain was calculated based on the utility function, U, described in the text. The welfare from savings bound (Savings) and the welfare bound based on minimum consumption (Floor) are also reported for each combination.

Consider the top row of Table 3—the row corresponding to actuarially fair markets. We see that as annuity markets become more flexible, moving from left to right on the chart, welfare increases. For example, as we move from an immediate annuity market to a longevity annuity market, welfare increases from \$30.4 to \$34.7. Moreover, the value of flexibility becomes more dramatic in the face of higher costs. For the money's worth value $K = 70\%$, as we move from an immediate annuity to a longevity annuity market the welfare increases from \$1.2 to \$20.6. Both welfare values decrease, as they must, but the percentage decrease of the longevity annuity is much less than that of the immediate annuity.

The second value in each welfare triplet is the welfare from savings. The actual welfare is the sum of the welfares from savings and consumption shifting, and we observe that former dominates the latter. Across all annuity enhancements and all cost levels, the savings welfare captures at least three-quarters of the actual welfare, and often captures ninety percent or more.[16] The welfare from shifting tends to be larger for longevity and zero-coupon annuity markets because these markets allow for more flexible spending patterns. The floor welfare, the third value in the triplet, is a bound on savings welfare. The quality of this bound depends critically on the utility function selected and the resulting taste for floor consumption. However, Table 3 does illustrate how, for this common class of utility specifications, welfare gains are reasonably well estimated by cost savings on floor consumption.

6. Completing the Annuity Market

In the preceding sections, we assumed the annuity market was characterized by bundling and frictions and investigated the types of annuity market enhancements likely to improve participation and generate welfare gains. In this section, we take a different approach. Here, we consider the types of annuity products demanded when annuity markets are complete. Characterizing annuity products that people want given access to a complete market could identify desirable annuity packages that effectively substitute for a complete market. To give an extreme example, suppose, when given access to infinite choice in the annuity market, everyone decides they want the same exact bundle of annuity payouts. This would be good news indeed for annuity providers. Instead of having to go to the expense of creating every possible annuity package, they need only

[16] While not reported in the text, this result held true across a variety of habit utility specifications. Additional data are available from the authors upon request.

create a single package. Access to that single package would provide all the benefits of a complete market.

To investigate this possibility, we assume our annuity market is complete in the sense that every state has an available survival-contingent asset. However, our annuity market is not frictionless. Annuity prices are described by equation (6). Since each state has a financial and an annuity asset available, we can define a state-specific spending improvement quotient, *q(s)*. Equation (6) and the definition of *q* imply:

$$(22) \quad q(s) = \frac{K_s - \pi_s}{\pi_s}$$

Financial and Annuity Market Separation Theorem

When an annuity market is incomplete, only specific annuity packages are available. In this case, costs are not the sole factor, and suitability depends on whether a package's payouts complement an investor's desired payouts. However, when an annuity market is complete, we can say a good deal about which annuities are purchased, independent of an investor's utility function. In particular, for both a rational and behavioral investor, there will exist a critical q_* that separates financial and annuity market purchases. Here, an optimal investment plan will only purchase financial assets in the states $s \in \Omega$ with a spending-improvement level *q(s)* less than q_* and only annuity assets in all states with a *q(s)* greater than the same q_* (see the Appendix for a demonstration of this result). We call this result the *Separation Theorem* for complete annuity markets.[17] Informally, annuities are purchased exclusively in high *q* states, whereas financial assets are exclusively purchased in low *q* states.

For a purely rational investor, q_* equals zero. In other words, in any state where the money's worth exceeds the survival probability the investor will rely on annuity assets for consumption. Similarly, in any state with a negative *q* the investor relies solely on financial market assets. This recovers Yaari's result of full annuitization since actuarially fair pricing implies all values of *q* are at least non-negative. It extends Yaari's analysis by describing the solution for complete annuity markets with frictions. Now,

[17] This generalizes the separation results presented in Scott et al. 2007

only states which offer a discount in the annuity market are supported with annuity investments. The result for a behavioral investor with a distaste of annuities is similar except now $q_* > 0$.

The Separation Theorem has significant implications for the annuity market. In general, optimal annuity bundles should never have payouts that occur in states with a survival probability too high to cover the costs of creating the annuity contract. For example, suppose an annuity market is characterized by a money's worth percentage of 85%. For our sample cohort, survival rates don't drop below 85% until age 77. Given access to a complete market, our investor would fund all consumption between the ages of 65 and 76 with financial market securities and all later life consumption with annuity market assets. In addition, if her preferences include a distaste for annuity investments, then the critical age would be even older.

The package of annuity products that all individuals want when markets are complete all share one common property. They all create sufficient time between the purchase date and the payout date to allow the survival discount to overcome any annuity costs. Surprisingly, this property characterizes few existing annuity products. With the exception of the recently-introduced longevity annuity, annuity packages uniformly include payouts with little to no potential for a survival discount. The alignment of the longevity annuity payouts with those desired given a complete market provide an important insight into why longevity annuities are a significant annuity market enhancement with the potential to increase participation and create welfare gains.

Conclusion

The disparity between economic prediction and observed annuity market demand has been well documented. However, the disparity between ideal and actual annuity markets is equally large. The difficulty associated with writing and monitoring a survival-contingent contract suggests annuity markets will never offer actuarially fair pricing nor will they likely approach financial markets in terms of flexibility. Given this reality, what types of annuity product innovations might close the gap between current and ideal annuity markets? In short, what makes a better annuity?

To answer this question, we must first identify the goal of annuity market innovation. We considered the annuity market implications of two separate objectives: increased participation and improved individual welfare. With increased participation as the objective, we suggested insurance companies focus on maximizing the spending improvement quotient, q. We demonstrated that annuity product introductions that do not alter the maximum q cannot influence participation. Similarly, introductions that significantly increase the maximum q could spur large increases in participation. While q is related to both mortality and costs, the most promising area for large increases in q is not lowering annuity costs, but rather offering annuity products focused on late-life payouts.

If our objective instead is to increase welfare, we suggested insurance companies pay close attention to the assets held by their potential customers. These assets offer important clues as to the types of payouts these individuals find desirable. In fact, the extent to which an annuity product innovation helps individuals replicate their existing portfolio payouts at a lower cost often provides a good estimate for total welfare gains. Thus, innovation focused on minimizing the cost of existing asset allocations likely garners a significant fraction of the potential welfare gain.

Finally, we considered which annuity packages are essential to ultimately completing the annuity market. Given complete annuity markets, annuity purchases should only occur when the mortality discount is sufficient to overcome any annuity costs. With positive costs, an annuity purchase thus requires a gap between the purchase date and the payout date. Surprisingly, the vast majority of annuity products available today bundle at least some payouts without a gap. This fundamental inefficiency suggests innovative new products may ultimately supplant many incumbent annuity products. While annuity demand is low, our research suggests that existing annuity markets offer ample opportunities for improvement. Hopefully, annuity innovations based on the guidelines identified in this research will ultimately result in the types of participation and welfare benefits Yaari identified over four decades ago.

References

Bengen, W. P. 1994. Determining Withdrawal Rates Using Historical Data. *Journal of Financial Planning*, 7(4): 171-180.

Brown, J.R. and M.J. Warshawsky. 2004. Longevity-Insured Retirement Distributions From Pension Plans: Market and Regulatory Issues. In Gale, W.G., Shoven, J.B., Warshawsky, M.J. (Eds.), *Public Policies and Private Pensions*. Brookings Institution Press, Washington, DC.

Brown, J.R., J.R. Kling, S. Mullainathan, and M.V. Wrobel. 2008. Why Don't People Insure Late Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle. *American Economic Review*, 98(2): 304-309.

Constantinides, G. M. 1990. Habit Formation: A Resolution of the Equity Premium Puzzle. *Journal of Political Economy* 98(3): 519-543.

Davidoff, T., J.R. Brown, and P.A. Diamond. 2005. Annuities and Individual Welfare. *American Economic Review* 95(5): 1573-1590.

Diamond, P. A., and J. A. Mirrlees. 2000. Adjusting One's Standard of Living: Two Period Models. In Hammond, P.J. and Myles, G. D. (Eds.), *Incentives, Organization, and Public Economics: Papers in Honour of Sir James Mirrlees*, Oxford University Press: 107-122.

Dus, I., R. Maurer, and O. S. Mitchell. 2005. Betting on Death and Capital Markets in Retirement: A Shortfall Risk Analysis of Life Annuities. *NBER Working Paper 11271.* National Bureau of Economic Research, Cambridge, MA.

Dushi, I. and A. Webb. 2004. Household Annuitization Decisions: Simulations and Empirical Analysis. *Journal of Pension Economics and Finance* 3(2): 109-143.

Horneff, W., R. Maurer, O.S. Mitchell, and I. Dus. 2006. Optimizing the Retirement Portfolio: Asset Allocation, Annuitization, and Risk Aversion. *NBER Working Paper 12392.* National Bureau of Economic Research, Cambridge, MA.

Hu, W. and J.S. Scott. 2007. Behavioral Obstacles in the Annuity Market. *Financial Analysts Journal*. 63(3): 71-82.

Kapur, S. and J.M. Orszag. 1999. A Portfolio Approach to Investment and Annuitization During Retirement. Mimeo, Birkbeck College, University of London, London, UK.

Luenberger, D. G. 1995. *Microeconomic Theory*, Appendix C, McGraw-Hill, New York, NY.: 458-463.

MetLife. 2004 Press Releases. MetLife Introduces "Longevity Insurance" to Help Protect Against Outliving Retirement Savings in Later years.

Milevsky, M.A. and V.R. Young. 2003. Annuitization and Asset Allocation. Working Paper, Individual Finance and Insurance Decisions Centre. University of Toronto, Toronto, Canada.

Mitchell, O.S., J.M. Poterba, M.J. Warshawsky, and J.R. Brown. 1999. New Evidence on the Money's Worth of Individual Annuities. *American Economic Review* 89(5): 1299-1318.

NAVA. 2008. *2008 Annuity Fact Book: A Guide to Information, Trends, and Data in the Annuity Industry.* NAVA, Inc. Reston, VA

Poterba, J.M. and M.J. Warshawsky. 2000. The Cost of Annuitizing Retirement Payouts from Individual Accounts. In Shoven, J.B. (Ed.), *Administrative Aspects of Investment-based Social Security Reform.* The University of Chicago Press, Chicago, IL.: 173-206.

Purcal, S. and J. Piggott. 2008. Explaining Low Annuity Demand: An Optimal Portfolio Application to Japan. *The Journal of Risk and Insurance* 75(2): 493-516.

Scott, J.S. 2008. The Longevity Annuity: An Annuity for Everyone? *Financial Analysts Journal*. 64(1): 40-48.

Scott, J.S., J.G. Watson and W. Hu. 2007. Efficient Annuitization: Optimal Strategies for Hedging Mortality Risk. Pension Research Council Working Paper 2007-9.

Scott, J.S., W.F. Sharpe and J.G. Watson. 2009. The 4% Rule: At What Price? *The Journal of Investment Management* 7(3): 1-18.

Sundaresan, S.M. 1989. Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth. *Review of Financial Studies* 2: 73-89.

Warshawsky, M. 1988. Private Annuity Markets in the United States: 1919-1984. *The Journal of Risk and Insurance*, 55(3): 518-528.

Yaari, M.E. 1965. Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *Review of Economic Studies* 32: 137-150.

Appendix

Rational investors will maximize their utility and choose the consumption stream given by the mathematical program described by equations (12). In this appendix, we consider the more general utility function V(c,α) which depends on the consumption stream c(s) for s ∈ Ω and allows for a potential dependency on the amount α invested in annuities.

$$(A1a)\quad V_* = V(c_*,\alpha_*) = \max V(c,\alpha)$$

$$(A1b)\quad c(s) = x(s) + \sum_{a\in A} y(a)\cdot G_a(s),\ s\in\Omega$$

$$(A1c)\quad \alpha = \sum_{a\in A} y(a)\cdot \Gamma_a$$

$$(A1d)\quad W_0 = \sum_{s\in\Omega} x(s)\cdot\psi(s) + \sum_{a\in A} y(a)\cdot\Gamma_a$$

$$(A1e)\quad x(s)\ge 0,\ s\in\Omega$$

$$(A1f)\quad y(a)\ge 0,\ a\in A$$

Here, the decision variables are c(s) and x(s) for s ∈ Ω, y(a) for a ∈ A, and α. The Karush-Kuhn-Tucker conditions (Luenberger 1995, eqs. C.4, 460-1) for this program simplify to the following set of equations:

$$(A2a)\quad c_*(s) = x_*(s) + \sum_{a\in A} y_*(a)\cdot G_a(s),\ s\in\Omega$$

$$(A2b)\quad \alpha_* = \sum_{a\in A} y_*(a)\cdot G_a(s)$$

$$(A2c)\quad W_0 = \sum_{s\in\Omega} x_*(s)\cdot\psi(s) + \sum_{a\in A} y_*(a)\cdot\Gamma_a$$

$$(A2d)\quad \mu(s) = \lambda_*\cdot\psi(s) - V_s(c_*,\alpha_*),\ s\in\Omega$$

$$(A2e)\quad x_*(s)\cdot\mu(s) = 0,\ x_*(s)\ge 0,\ \mu(s)\ge 0,\ s\in\Omega$$

$$(A2f)\quad v(a) = \Gamma_a\cdot\left[\lambda_* - V_\alpha(c_*,\alpha_*)\right] - \sum_{s\in\Omega} V_s(c_*,\alpha_*)\cdot G_a(s),\ a\in A$$

$$(A2g)\quad y_*(a)\cdot v(a) = 0,\ y_*(a)\ge 0,\ v(a)\ge 0,\ a\in A$$

We assume that V is concave and that these conditions are both necessary and sufficient for a maximum. Further, we assume that the budget constraint, equation (A2c), is binding and that its Lagrange multiplier λ is positive.

We use equations (A2) to demonstrate a number of useful theorems. Before proceeding, we derive an alternative form of equation (A2f) that we use extensively in these proofs. If we eliminate the partial derivates V_s from equation (A2f) using equation (A2d), we obtain:

$$(A3a) \quad \upsilon(a) = \sum_{s\in\Omega} \mu(s) \cdot G_a(s) - \lambda \cdot \Gamma_a \cdot \left[q(a) - P_*\right]$$

In the above equation, we have introduced the parameter q(a), an annuity's spending-improvement quotient:

$$(A3b) \quad q(a) \equiv \left(\psi_{G_a} - \Gamma_a\right)/\Gamma_a$$

$$(A3c) \quad \psi_{G_a} \equiv \sum_{s\in\Omega} \psi(s) \cdot G_a(s)$$

The price ψ_G is the price of replicating the annuity's payouts with financial assets. Also in equation (A3a), we have introduced P_*, the penalty for purchasing annuities:

$$(A3d) \quad P_* \equiv -V_\alpha(c_*, \alpha_*)/\lambda_*$$

Since we assume that utility decreases as the wealth invested in annuities increases, the penalty P_* is positive for behavioral investors. For purely rational investors, this parameter is zero.

Theorem: The spending-improvement quotient of any purchased annuity must be greater than or equal to the behavioral penalty, a non-negative number.

Proof: Suppose that $y(a) > 0$ for some annuity $a \in A$. It follows from the complimentary slackness condition equation (A2g) that $\nu(a)$ is zero, and hence the right side of equation (A3a) is zero:

$$(A4a) \quad \lambda \cdot \Gamma_a \cdot \left[q(a) - P_*\right] = \sum_{s\in\Omega} \mu(s) \cdot G_a(s)$$

We have $\mu(s) \geq 0$ by equation (A2e), and since $G_a(s) \geq 0$, the right side of the above equation is non-negative. Further, since λ is positive, we must have $q(a) \geq P_*$. QED

Corollary: If an annuity is purchased, then its price is no greater than the price of the bundle of financial assets that replicates its payouts.

Proof: From the previous theorem, the spending-improvement quotient q(a) must be non-negative. It follows from its definition that if $q \geq 0$, then $\psi_G \geq \Gamma_a$. QED

Theorem: If the spending-improvement quotient of any annuity is greater than the behavioral penalty, then at least one annuity product is purchased.

Proof: For any annuity product a, the Lagrange multiplier for its no short-sale constraint must be non-negative, i.e., $\nu(a) \geq 0$. Using equation (A3a), we get the following inequality for every $a \in A$:

$$(A5a) \quad \sum_{s \in \Omega} \mu(s) \cdot G_a(s) \geq \lambda \cdot \Gamma_a \cdot \left[q(a) - P_* \right]$$

If for some annuity a, its quotient q(a) exceeds P_*, then the sum on the left side of equation (A5a) is strictly positive. Hence, there must be at least one state $s' \in \Omega$ where both $G_a(s')$ and $\mu(s')$ are simultaneously greater than zero. However, if $\mu(s') > 0$, then it follows from equation (A2e) that $x(s') = 0$. We know from our no-starvation assumption that every state must receive either a financial or annuity payout. Since the payout in state s' is not financial, it must be from an annuity, say a'. Though a' and a are not necessarily the same annuity, they will have at least one payout state in common. QED

Corollary: For non-behavioral investors, if the price of any annuity is strictly less than the price of the bundle of financial assets that replicates its payouts, then at least one annuity product is purchased.

This theorem follows from the previous theorem when P_* is zero. QED

Separation Theorem: When the annuity market is complete, annuity state-claims are purchased on the states for which $q(s) > P_*$, and financial state claims are purchased on the states for which $q(s) < P_*$.

Proof: For a complete annuity market, we have an annuity state-claim corresponding to each financial state-claim, i.e., $A \equiv S$. Further, for each annuity state-claim a, the payout function $G_a(s)$ equals one when s equals a, and equals zero otherwise. The sum in equation (A3a) simplifies, and the equation reduces to:

$$(A6) \quad \lambda \cdot \Gamma_s \cdot [q(s) - P_*] = \mu(s) - \upsilon(s)$$

When q(s) exceeds P_*, $\mu(s)$ must be positive, and from equation (A2e), it follows that $x_*(s) = 0$, i.e., no financial state-claims are purchased. Since we do not allow starvation states, we must have $y_*(s) > 0$. Similarly, we can show that $q(s) < P_*$, implies $x_*(s) > 0$ and $y_*(s) = 0$. QED

Theorem: For an investor with all her wealth in financial assets, the marginal utility of switching money out of financial assets and into annuity assets is proportional to $Q - P_*$, where Q is the market's spending-improvement quotient and P_* is her behavioral penalty.

An investor who chooses not to purchase annuities will solve equations (A1), but with the no-short constraints ($y \geq 0$) changed to no-purchase constraints ($y = 0$). In this case, equations (A2) remain valid, except that we drop the non-negativity constraints on the Lagrange multipliers ν—they now can have either sign. If we were to relax one of the no-purchase constraints, then the marginal increase in maximum utility for an increase in shares of this annuity is proportional to the constraint's multiplier, i.e., $dV_*/dy = -\nu$, where ν is given by equation (A3a). Note that if ν is positive, then there is no advantage to purchasing this annuity. Since all consumption is funded with financial assets, we have $x(s) > 0$ and $\mu(s) = 0$ for all $s \in \Omega$. In this case, the marginal utility at $y = 0$ simplifies is:

$$(A7a) \quad dV_* / dy\big|_{y=0} = \lambda_* \cdot \Gamma_a \cdot [q(a) - P_*]$$

Alternatively, we can compute the marginal utility of investing an amount α in this same annuity, and since the investment equals the product of shares and price, we have:

$$(A7b) \quad dV_* / d\alpha\big|_{\alpha=0} = \lambda_* \cdot [q(a) - P_*]$$

Equations (A7) are valid for relaxing the no-purchase constraint on any one annuity, while enforcing it on all the remaining products. If we simultaneously open the contest to all annuities, the product(s) with the largest q will win the investment, and so for an unrestricted investment α, we have:

$$(A8a) \quad dV_* / d\alpha\big|_{\alpha=0} = \lambda_* \cdot \left(Q - P_*\right)$$

$$(A8b) \quad Q = \max_{a \in A} q(a)$$

In equations (A8), the market's spending-improvement quotient Q is the largest quotient of any of its products. QED