

Annuities and Individual Welfare

March 17, 2005

Abstract: Advancing annuity demand theory, we present sufficient conditions for the optimality of full annuitization under market completeness that are substantially less restrictive than those used by Yaari (1965). We examine demand with market incompleteness, finding that positive annuitization remains optimal widely, but complete annuitization does not. How uninsured medical expenses affect demand for illiquid annuities depends critically on the timing of the risk. A new set of calculations with optimal consumption trajectories very different from available annuity income streams still shows a preference for considerable annuitization, suggesting that limited annuity purchases are plausibly due to psychological or behavioral biases. (JEL D11, D91, E21, H55, J14, J26)

Since the seminal contribution of Yaari (1965) on the theory of a life-cycle consumer with an unknown date of death, annuities have played a central role in economic theory. His widely cited result is that certain consumers should fully annuitize all of their savings. However, these consumers were assumed to satisfy several very restrictive assumptions: they were von Neumann-Morgenstern expected utility maximizers with intertemporally separable utility, they faced no uncertainty other than time of death, they had no bequest motive, and the annuities available for purchase were actuarially fair. While the subsequent literature on annuities has occasionally relaxed one or two of these assumptions, the “industry standard” is to maintain most of these conditions. In particular, the literature has universally retained expected utility and additive separability, the latter dubbed “not a very happy assumption” by Yaari. While Yaari (1965) and Bernheim (1987a,1987b) provide intuitive explanations of why the Yaari result may not depend on these strict assumptions, the generality of this result has not been formally shown in the literature.

The first contribution of this paper is to present sufficient conditions substantially weaker than those imposed by Yaari under which full annuitization is optimal. The heart of the argument can be seen by comparing a one-year bank certificate of deposit (CD), paying an interest rate r to a security that pays a higher interest rate at the end of the year conditional on living, but pays nothing if you die before year-end. If you attach no value to wealth after death, then the second, annuitized, alternative is a dominant asset. This simple comparison of otherwise matching assets, when articulated in a setting that confirms its relevance, lies behind the Yaari result. The dominance comparison of matching assets is not sensitive to their financial details, but does rely

on the identical liquidity in the annuitized and non-annuitized assets.

This paper explores the implications of this comparison for asset demand in different settings, including both Arrow-Debreu complete markets and incomplete market settings. In the Arrow-Debreu complete market setting, sufficient conditions for full annuitization to be optimal are that consumers have no bequest motive and that annuities pay a rate of return to surviving investors, net of administrative costs, that is greater than the return on conventional assets of matching financial risk. Thus, when markets are complete, full annuitization is optimal without assuming exponential discounting, the expected utility axioms, intertemporal separability, or actuarially fair annuities. We also relate the size of the welfare gain to the trajectory of optimal consumption.

A second contribution is to examine annuity demand in some incomplete market settings. If some desired consumption paths are not available when all wealth is annuitized in an incomplete annuity market, full annuitization may no longer be optimal. For example, if an individual desires a steeply downward sloping consumption path, but only constant real annuities are available, then full annuitization is no longer optimal. Thus, we explore conditions for the optimum to include partial annuitization. We also consider partial annuitization with a bequest motive.

Another example involves the relative liquidity of annuities and bonds in the absence of insurance for medical expenditure shocks. The effect on annuity demand depends on the timing of the risk. An uninsurable risk early in life may reduce the value of annuities if it is not possible to sell or borrow against future payments of the fixed annuity stream but it is possible to do so with bonds. In contrast, loss of insurability of a shock occurring later in life may increase annuity purchases as a substitute for medical insurance that has an inherent annuity character. Thus, an annuity can be a better substitute than a bond for long-term care insurance.

Most practical questions about annuitization (e.g., the appropriate role of annuities in public pension systems) are concerned with partial annuitization. The general theory itself is insufficient to answer questions about the optimal fraction of annuitized wealth, and thus a large simulation literature has developed. Our third contribution is to extend the simulation literature by creating a new “stress test” of annuity valuation. We do so by generating optimal consumption trajectories that differ substantially from what is offered by a fixed real annuity contract, and showing that even under these highly unfavorable conditions, the majority of wealth is still optimally annuitized. To generate the unfavorable match between optimal consumption and the annuity trajectory, we

allow a person’s utility to depend on how present consumption compares to a standard of living to which the individual has become accustomed, which is itself a function of past consumption. We model this “internal habit” as in Diamond and Mirrlees (2000).¹

These results imply that annuities are quite valuable to utility maximizing consumers, even under conditions that result in a very unfavorable mismatch between available annuity income streams and one’s desired consumption path. Thus, while incomplete markets, when combined with preferences for consumption paths that deviate substantially from those offered by current annuity products, can certainly explain the lack of *full* annuitization, it is difficult to explain the near universal lack of *any* annuitization outside of Social Security and defined benefit pensions plans, at least at the higher end of the wealth distribution where Social Security is a small part of one’s portfolio and SSI is not relevant. This finding is strongly reminiscent of the literature on life insurance, which is a closely related product.² The literature on life insurance has documented a severe mismatch between life insurance holdings of most households and their underlying financial vulnerabilities (see e.g. Bernheim, Forni, Gokhale and Kotlikoff (2003); Auerbach and Kotlikoff (1987), Auerbach and Kotlikoff (1991)). These papers, taken together, are suggestive of psychological or behavioral considerations at play in the market for life-contingent products that have not yet been incorporated into standard economic models.

The focus of the paper is on the properties of annuity demand in different market settings. We do not address the more complex equilibrium question of what determines the set of annuity products in the market. In light of the value of annuities to consumers in standard models, we think examination of equilibrium would have to include the supply response to demand behavior that is not consistent with standard utility maximization. We do not develop such a behavioral theory of annuity demand, but rather clarify the mismatch between observed demand behavior and the value of annuitization in a standard utility maximization setting.

The paper proceeds as follows: In section I, we provide a general set of sufficient conditions under which full annuitization is optimal under complete markets. Section II discusses of why

¹Different models of intertemporal dependence in utility are discussed in, for example, Dusenberry (1949), Abel (1990), Constantinides (1990), Deaton (1991), Campbell and Cochrane (1999), Campbell (2002) and Gomes and Michaelides (2003).

²As discussed by Yaari (1965) and Bernheim (1991), the purchase of a pure life insurance policy can be viewed as the selling of an annuity.

full annuitization may no longer be optimal with incomplete markets, paying particular attention to the role of illiquidity of annuities. Section III briefly discusses a bequest motive. In Section IV we report simulation results reflecting the quantitative importance of market incompleteness with habit formation, showing that even when the liquidity constraints on annuities are binding, individuals still prefer a high level of annuitization. Section V concludes.

I. Complete Markets

Much of the focus of the annuities literature has been an attempt to reconcile Yaari's (1965) "full annuitization" result with the empirical fact that few people voluntarily annuitize any of their private savings.³ This issue is of theoretical interest because it bears upon the issue of how to model consumer behavior in the presence of uncertainty. It is also of policy interest because of the shift in the US from defined benefit plans, which typically pay out as an annuity, to defined contribution plans that rarely offer retirees directly the opportunity to annuitize. Annuitization is also important in the debate about publicly provided defined contribution plans.

This section derives a general set of conditions under which full annuitization is optimal, relaxing many of the assumptions in the original Yaari formulation. We begin with a simple two period model, and then show formally the generalization to many periods and many states.

A. *The Optimality of Full Annuitization in a Two Period Model with No Aggregate Uncertainty*

Analysis of intertemporal choice is greatly simplified if resource allocation decisions are made completely and all at once, that is, without additional, later trades. Consumers will be willing to commit to a fixed plan of expenditures if, at the start of time, they are able to trade goods across all time and all states of nature, as is standard in the complete market Arrow-Debreu model.

Yaari considered annuitization in a continuous time setting where consumers are uncertain only about the date of death. Some results, however, can be seen more simply by dividing time into two discrete periods: the present, period 1, when the consumer is definitely alive and period 2, when

³This assertion is consistent with the large market for what are called variable annuities since these insurance products do not include a commitment to annuitize accumulations, nor does there appear to be much voluntary annuitization. See for example Brown and Warshawsky (2001).

the consumer is alive with probability $1 - q$.⁴

By writing $U = U(c_1, c_2)$, we allow for a very general formulation of utility in a two-period setting with the assumptions that there is no bequest motive and that only survival to period 2 is uncertain. Lifetime utility is defined over first period consumption, c_1 , and consumption in the event that the consumer is alive in period 2, c_2 . We drop the requirement of intertemporal separability, allowing for the possibility that the utility from second-period consumption may depend on the level of first period consumption. Additionally, this formulation does not require that preferences satisfy the axioms for U to be an expected value.

The optimal consumption decision and the welfare evaluation of annuities can be determined using a dual approach: minimizing expenditures subject to attaining at least a given level of utility. We measure expenditures in units of first period consumption. Assume that there are two securities available. The first is a bond that returns R_B units of consumption in period 2, whether the consumer is alive or not, in exchange for each unit of the consumption good in period 1. The second is an annuity that returns R_A in period 2 if the consumer is alive and nothing otherwise.

An actuarially fair annuity would yield $R_A = R_B/(1 - q)$. Adverse selection and higher transaction costs for paying annuities than for paying bonds may drive returns below this level. However, because any consumer will have a positive probability of dying between now and any future period, thereby relieving borrowers' obligation, we make the weak assumption that $R_A > R_B$.⁵

If we denote by A savings in the form of annuities and by B savings in the form of bonds, and if there is no other income in period 2 (e.g., the individual is retired), then $c_2 = R_A A + R_B B$, and expenditures for lifetime consumption are simply $E = c_1 + A + B$. Thus, the expenditure minimization problem can be written as:

$$\min_{c_1, A, B} c_1 + A + B \text{ s.t. } U(c_1, R_A A + R_B B) \geq \bar{U}. \quad (1)$$

We further impose the constraint that $B \geq 0$, i.e., that the individual not be permitted to die in debt. Otherwise with $R_A > R_B$, purchasing annuities and selling bonds in equal numbers would

⁴A two period model with a single consumer good in each dated event precludes trade after the first period, but in a complete market setting, this is irrelevant.

⁵That $R_B < R_A < R_B/(1 - q)$ is supported empirically by Mitchell et al (1999). If the first inequality were violated, annuities would be dominated by bonds.

cost nothing and yield positive consumption when alive in period 2, but leave a debt if dead, leaving lenders with expected financial losses in total.

This setup leads immediately to the optimality of full annuitization. If $B > 0$, then one is able to reduce expenditures, while holding the consumption vector fixed, by selling R_A/R_B of the bond and purchasing one unit of the annuity (noting that $R_A > R_B$). Thus, the solution to this expenditure minimization problem is to set $B = 0$, i.e., to annuitize one's wealth fully. The intuition is that allowing individuals to substitute annuities for conventional assets yields an arbitrage-like gain when the individual places no value on wealth when not alive. Such a gain enhances welfare independent of assumptions about preferences beyond the lack of utility from a bequest. Nor must annuities be actuarially fair. Indeed, all that is required is for consumers to have no bequest motive and for the payouts from the annuity to exceed that of conventional assets for the survivor.

Equation (1) also indicates an approach to evaluating the gain from an increased opportunity to annuitize. Consider the minimization under the further constraint of an upper bound on purchases of annuities, $A \leq \bar{A}$. We know that utility-maximizing consumers will take advantage of an arbitrage-like opportunity to annuitize as long as bond holdings are positive and can therefore be used to finance the purchase. With no annuities available, bond holdings will be positive if second-period consumption is positive, as is ensured by the plausible condition that zero consumption is extremely bad:

ASSUMPTION 1:

$$\lim_{c_t \rightarrow 0} \partial U / \partial c_t = \infty : t = 1, 2$$

Allowing consumers previously unable to annuitize any wealth to place a small amount of their savings into annuities (increasing A from zero) leaves second period consumption unchanged (since the cost of the marginal second-period consumption is unchanged, so too, is the optimal level of consumption in both periods). Thus a small increase in \bar{A} from zero reduces the cost of achieving a given level of utility by $1 - (R_A)/(R_B) < 0$. This is the welfare gain from increasing the limit on available annuities for an optimizing consumer with positive bond holdings.⁶

If the upper bound constraint on available annuities is large enough that bond holdings are zero, then the price of marginal second period consumption (up to \bar{A}) falls from $1/(R_B)$ to $1/(R_A)$. With

⁶This point is made in the context of time-separable preferences by Bernheim (1987b).

a fall in the cost of marginal second-period consumption, its compensated level will rise. Thus the welfare gain from unlimited annuity availability is made up of two parts. One part is the savings while financing the same consumption bundle as when there is no annuitization, and the second is the savings from adapting the consumption bundle to the change in prices. We can measure the welfare gain in going from no annuities to potentially unlimited annuities from the expenditure function defined over the price vector $(1, p_2)$ and the utility level \bar{U} . Integrating the derivative of the expenditure function evaluated at the two prices $1/R_A$ and $1/R_B$:

$$E|_{\bar{A}=0} - E|_{\bar{A}=\infty} = E(1, 1/R_B, \bar{U}) - E(1, 1/R_A, \bar{U}) = - \int_{1/R_A}^{1/R_B} c_2(p_2) dp_2, \quad (2)$$

where $c_2(p_2)$ is compensated demand arising from minimization of expenditures equal to $c_1 + c_2 p_2$ subject to the utility constraint without a distinction between asset types. Consumers who save more (have larger second-period consumption) benefit more from the ability to annuitize completely.

B. *The Optimality of Full Annuitization with Many Periods and Many States*

While a two-period model with no uncertainty other than length of life has a clarity that derives from its simplicity, real consumers face a more complicated decision setting. In particular, they face many periods of potential consumption and each period may have several possible states of nature. For example, a 65 year-old consumer has some probability of surviving to be a healthy and active 80 year-old, some chance of finding herself sick and in a nursing home at age 80, and some chance of not being alive at all at age 80. Moreover, returns on some assets are stochastic. In this section, we show that the optimality of complete annuitization survives subdivision of the aggregated future defined by c_2 into many future periods and states, as long as markets are complete.

A simple subdivision would be to add a third period, while continuing with the assumption of no other uncertainty. In keeping with the complete market setting, we have bonds and annuities that pay out separately in period 2 with rates R_{B2} and R_{A2} , and period 3 with rates R_{B3} and

R_{A3} .⁷ That is, defining bonds and annuities purchased in period 1 with the appropriate subscript:

$$E = c_1 + A_2 + A_3 + B_2 + B_3 \quad (3)$$

$$c_2 = R_{B2}B_2 + R_{A2}A_2 \quad (4)$$

$$c_3 = R_{B3}B_3 + R_{A3}A_3 \quad (5)$$

If our assumption that the return on annuities exceeds that of bonds holds period by period, then our full optimization result extends trivially. Note that the standard definition of an Arrow security distinguishes between states when an individual is alive and when he or she is not alive. That is, a standard Arrow security is an “Arrow annuity.” We have set up what we call “Arrow bonds” (here B_2 and B_3) by combining matched pairs of events that differ in whether the consumer’s death has occurred. This representative of a standard bond is what becomes a dominated asset once one can separate the Arrow bond into two separate Arrow securities, and the consumer can choose to purchase only one of them. In a setting with additional sources of uncertainty, the combination of a matched pair of events to depict a non-annuitized asset is straightforward when the death of the particular consumer is independent of other events and unimportant in determining equilibrium. In some settings there may not be such a decomposition of a bond because the recognition of important differences between different events is strongly related to the survival of the individual.⁸

To take the next logical step, and assuming we can decompose Arrow bonds, we continue to treat c_1 as a scalar and interpret c_2 , B_2 and A_2 as vectors with entries corresponding to arbitrarily many future periods ($t \leq T$), within arbitrarily many states of nature ($\omega \leq \Omega$). R_{A2} (R_{B2}) is then a matrix with columns corresponding to annuities (bonds) and rows corresponding to payouts by period and state of nature. Thus, the assumption of no aggregate uncertainty can be dropped. Multiple states of nature might refer to uncertainty about aggregate issues such as output, or individual specific issues beyond mortality such as health. To extend the analysis, we assume that the consumer is sufficiently “small” (and with death uncorrelated with other events) that for each state of nature where the consumer is alive, there exists a state where the consumer is dead and

⁷In keeping with the complete market setting, there is no arrival of asymmetric information about future life expectancy.

⁸If the death of a consumer only occurs with other large changes, then there is no way to construct an Arrow bond that differs from an Arrow annuity only in the death of a single consumer, for example, if an individual will die in some future period if and only if an influenza epidemic occurs.

the equilibrium relative prices are otherwise identical. Completeness of markets still has Arrow bonds that represent the combination of two Arrow securities. Note, however, that the ability to construct such bonds is not necessary for the result that a consumer without a bequest motive does not purchase consumption when not alive in the complete market setting.

Annuities paying in only one dated event are contrary to conventional life annuities that pay out in every year until death. However, with complete markets, separate annuities with payouts in each year can be combined to create such conventional annuities.

It is clear that the analysis of the two-period model extends to this setting, provided we maintain the standard Arrow-Debreu market structure and assumptions that do not allow an individual to die in debt since the consumer can purchase a combination of annuities with a structure of benefits across time and states as desired. In addition to the description of the optimum, the formula for the gain from allowing more annuitization holds for state-by-state increases in the level of allowed level of annuitization. Moreover, by choosing any particular price path from the prices inherent in bonds to the prices inherent in annuities, we can measure the gain in going from no annuitization to full annuitization. This parallels the evaluation of the price changes brought about by a lumpy investment (see Diamond and McFadden (1974)). Hence, with complete markets, preferences only matter through optimal consumption; this fact may clarify, for example, the *unimportance* of additive separability to the result of complete annuitization.

Stating the result more formally, the full annuitization result is a corollary of the following:

THEOREM 1: If there is no future trade once portfolio decisions have been made, if there is no bequest motive and if there is a set of annuities with payouts per unit of investment that dominate the payouts of some subset of bonds that are held, then a welfare gain is available by selling the subset of bonds and replacing the bonds with the annuities, so long as this transfer does not lead to negative wealth in any state.

COROLLARY 1: In a complete market Arrow-Debreu equilibrium, a consumer without a bequest motive annuitizes all savings.

Thus, with complete markets, *the result of full annuitization extends to many periods, the presence of aggregate uncertainty, actuarially unfair but positive annuity premiums, and intertemporally*

dependent utility that need not satisfy the expected utility axioms. This generalization of Yaari holds so long as markets are complete. Thus, if the puzzle of why so few individuals voluntarily annuitize is to be solved within a rational, life-cycle framework, we have shown that the answer does not lie in the specification of the utility function per se (beyond the issue of a bequest motive). We briefly consider the role of annuitization with a bequest motive below. However, the results do demonstrate the importance of market completeness, an issue that we turn to next.

II. Incomplete Markets

With complete markets, a higher yield of an Arrow annuity over the matching Arrow bond, dated event by dated event, leads directly to the result that a consumer without a bequest motive fully annuitizes. But markets are not complete. There are two forms of incompleteness that we consider. The first is that the set of annuities is highly limited, relative to the set of non-annuitized securities that exist.⁹ The second is the incompleteness of the securities market.

A. Incomplete Annuity Markets

Most real world annuity markets require that a consumer purchase a particular time path of payouts, thereby combining in a single security a particular “compound” combination of Arrow annuities. Privately purchased immediate life annuities are usually fixed in nominal terms, or offer a predetermined nominal slope such as a 5 percent increase per year. Variable annuities link the payout to the performance of a particular underlying portfolio of assets and combine Arrow securities in that way. CREF annuities are also participating, which means that the payout also varies with the actual mortality experience for the class of investors. To explore issues raised by such restrictions, we restrict our analysis to a single kind of annuity - a constant real annuity - although we state some of our results in a more general vocabulary. We examine the demand for such an annuity, distinguishing whether all trade must occur at a single time or whether it is possible to also purchase bonds later.

Trade Occurs All at Once

⁹Explaining why this is the case would take us into the industrial organization of insurance supply, which would necessarily make use of consumer understanding and perceptions of insurance. These issues are well beyond the scope of this paper. Rather, we analyze rational annuity demand in different market settings that are taken as given.

For the case of a conventional annuity, we must revise the superior return condition for Arrow annuities that $R_{At\omega} > R_{Bt\omega} \forall t\omega$. An appropriate formulation for a compound security is that it cost less than purchasing the same consumption vector using bonds. Define by ℓ a row vector of ones with length equal to the number of states of nature occurring in the annuity and so distinguished by bonds. Let the set of bonds continue to be represented by a vector with elements corresponding to the columns of the matrix of returns R_B and let R_A be a vector of annuity payouts multiplying the scalar A to define state-by-state payouts. Then the cost of the bonds exceeds the cost of the annuity under the assumption:

ASSUMPTION 2: *For any annuitized asset A and any collection of conventional assets B , $R_A A = R_B B \Rightarrow A < \ell B$.*

For example, if there is an annuity that costs one unit of first period consumption per unit and pays R_{A2} per unit of annuity in the second period and R_{A3} per unit of annuity in the third period, then we would have $1 < R_{A2}/R_{B2} + R_{A3}/R_{B3}$.¹⁰ By linearity of expenditures, this implies that any consumption vector that may be purchased strictly through annuities is less expensive when financed through annuities than when purchased by a set of bonds with matching payoffs.

Consider a three-period model, with complete bonds and a single available annuity and no opportunity for trade after the initial contracting. The minimization problem is now

$$\min_{c_1, A, B} : c_1 + B_2 + B_3 + A \tag{6}$$

$$\text{s.t.} : U(c_1, R_{B2}B_2 + R_{A2}A, R_{B3}B_3 + R_{A3}A) \geq \bar{U} \tag{7}$$

$$B_2 \geq 0, B_3 \geq 0 \tag{8}$$

Given our return assumption and positive consumption whenever alive, then, we have an arbitrage-like dominance of the annuity over the matching combination of bonds as long as this trade is feasible. Thus we can conclude that some annuitization is optimal and that the optimum has zero bonds in at least one dated event. The logic extends to a setting with more dates and states. However we would not get complete annuitization if the consumption pattern with complete annuitization is worth changing by purchasing a bond. That is, purchasing a bond would be

¹⁰The right hand side represents the required investments in two Arrow bonds that cost one unit each in the first period to replicate the annuity payout.

worthwhile if it raises utility by more than the decline from decreased first period consumption. Thus, denoting partial derivatives of the utility function with subscripts, there will be positive bond holdings if we satisfy either of the conditions:

$$U_1(c_1, R_{A2}A, R_{A3}A) < R_{B2}U_2(c_1, R_{A2}A, R_{A3}A) \quad (9)$$

or

$$U_1(c_1, R_{A2}A, R_{A3}A) < R_{B3}U_3(c_1, R_{A2}A, R_{A3}A). \quad (10)$$

By our return assumption, we can not satisfy both of these conditions at the same time, but we might satisfy one of them.

Additional Trading Opportunities

The previous analysis stayed with the setting of a single time to trade. If there are additional trading opportunities, the incompleteness of annuity markets may mean that such opportunities are taken. Staying within the setting of perfectly predicted future prices and assuming that prices for the same commodity purchased at different dates are all consistent, we can see that repeated bond purchase may increase the degree of annuitization. If desired consumption occurs later than with consuming all of the annuitized benefit, then the ability to save by purchasing bonds, rather than consuming all of the annuity payment, means that annuitization is made more attractive.

Returning to the three-period model with only mortality uncertainty, we can write this by denoting saving at the end of the second period by Z ($Z \geq 0$). We assume that the return on savings between the second and third periods Z is consistent with the other bond returns ($R_Z = R_{B3}/R_{B2}$). The minimization is now:

$$\min_{c_1, A, B, Z} c_1 + B_2 + B_3 + A \quad (11)$$

$$s.t. : U(c_1, R_{B2}B_2 + R_{A2}A - Z, R_{B3}B_3 + R_{A3}A + (R_{B3}/R_{B2})Z) \geq \bar{U}. \quad (12)$$

The restriction of not dying in debt is the non-negativity of wealth (including the value of future payments) if A equals zero:

$$R_{B2}B_2 + (R_{B2}/R_{B3})R_{B3}B_3 \geq 0 \quad (13)$$

$$R_{B3}B_3 + (R_{B3}/R_{B2})Z \geq 0. \quad (14)$$

Dissaving after full annuitization (if possible) would not be attractive if:

$$R_{B2}U_2(c_1, R_{A2}A, R_{A3}A) \leq R_{B3}U_3(c_1, R_{A2}A, R_{A3}A). \quad (15)$$

Under Assumption 2, (15) is now sufficient for the result of full annuitization of initial savings. To see this, note that Assumption 2 implies that $R_{B3} < R_{A2}R_Z + R_{A3}$. Thus, holdings of B_3 are dominated by the annuity with the second period return fully saved. Positive holdings of B_2 are ruled out by Assumption 2 and condition (15) since they imply:

$$R_{B2}U_2 < R_{A2}U_2 + (R_{A3}/R_{B3})R_{B2}U_2 \leq R_{A2}U_2 + R_{A3}U_3, \quad (16)$$

which is inconsistent with the FOC for positive holdings of both A and B_2 (and some annuitization is part of the optimum). In the commonly used model of intertemporally additive preferences with identical period utility functions, a constant discount rate and a constant interest rate, a sufficient condition for full initial annuitization in a constant real annuity is thus that $\delta(1+r) \geq 1$. If the available annuity (a constant real benefit) provides consumption later than an individual wants, then the assumed illiquidity of an annuity limits its attraction. We consider illiquidity below.

B. *Incomplete Securities and Annuity Markets: The Role of Liquidity*

A widely recognized basis for incomplete annuitization is that there may be an expenditure need in the future which cannot be insured. This might be an individual need, like a medical expense that is not insurable, or an aggregate event such as unexpected inflation, which lowers the real value of nominal annuity payments in the absence of real annuities. With incomplete markets, the arbitrage-like dominance argument used above will no longer hold if bonds are liquid while annuities are not.¹¹ We assume total illiquidity of annuities without exploring the possible arrival of asymmetric information about life expectancy which would naturally reduce the liquidity of annuities far more than the liquidity of non-traded bonds such as certificates of deposit.

We show, however, that the presence of uninsured risks may add to or subtract from the optimal fraction of savings annuitized, depending on the nature of the risk. We illustrate the role of illiquidity by examining two cases: uninsured medical expenditures, and the arrival of asymmetric information combined with inferior annuity returns.

¹¹More generally, it is well known that lifecycle consumers may be unwilling to invest in illiquid assets when they face stochastic cash needs Huang (2003).

Annuities and Medical Expenditures

For concreteness, let us start with the two period model above, where the only uncertainty is length of life. To this model let us add a risk of a necessary medical expense which we consider separately for each period. Assume that this expenditure enters into the budget constraint as a required expenditure, but does not enter into the utility function. Assume further that the occurrence of illness has no effect on life expectancy.

If it is possible to fully insure future medical expenses on an actuarially fair basis then the optimal plan is full medical insurance and full annuitization. Thus removing the ability to insure medical expenses, while continuing to assume that annuity benefits can be purchased separately period-by-period, can only raise the demand for bonds and may do so if the medical risk occurs in period 1, but not if it occurs in period 2. That is, the illiquidity of annuities may be relevant if the risk occurs early in life, but not toward the end of life. For example, contrast the cost of a hospitalization early in retirement with the need for a nursing home toward the end of life. The former calls for shifting expenses to earlier and so increases the value of an asset that permits such a change. Since medical expenses only occur for the living, annuities are a better substitute for nonexistent insurance for medical expenses later in life than are bonds.

To the two-period problem considered in (1), we add the risk of a first-period medical expense of size, M , with probability m and insurance at cost I , paying a benefit of β per dollar of insurance. If there is no additional trading after the initial purchases, the problem is:

$$\min_{c_1, A, B, I} c_1 + A + B + I \quad (17)$$

$$\text{s.t. } (1 - m)U(c_1, R_AA + R_BB) + mU(c_1 - M + \beta I, R_AA + R_BB) \geq \bar{U}. \quad (18)$$

In the case of no trading after the initial date, annuities continue to dominate bonds. If the insurance is actuarially fair, there will be complete medical insurance (and so c_1 , A , and B are the same as if there were no risk and expenditures were reduced by mM). With less favorable medical insurance, the presence of both risk and insurance generally affects the level of illiquid savings and so the level of annuitization, but all savings are annuitized since bonds are still dominated.

To bring out the role of liquidity, assume that bonds can be sold in the first period with an early redemption penalty, but annuities can not be sold. The analysis would be similar with liquid annuities that had a larger penalty for early redemption or the ability to reduce spending and add

to bond holdings (at a lower interest rate) after a realization of no health risk. Then the problem becomes:

$$\min_{c_1, A, B, I, Z} c_1 + A + B + I \quad (19)$$

$$\text{s.t. } (1 - m)U(c_1, R_A A + R_B B) + mU(c_1 - M + \beta I + \alpha_B Z, R_A A + R_B (B - Z)) \geq \bar{U}. \quad (20)$$

where Z is the value of bonds withdrawn early and α_B ($\alpha_B < 1$) is the fraction of value received net of the early withdrawal penalty.

Since annuities still dominate any bonds that would never be cashed in, the level of bond holdings would not exceed the amount cashed in early and we can rewrite the problem as

$$\min_{c_1, A, B, I} c_1 + A + B + I \quad (21)$$

$$\text{s.t. } (1 - m)U(c_1, R_A A + R_B B) + mU(c_1 - M + \beta I + \alpha_B B, R_A A) \geq \bar{U}. \quad (22)$$

In this case, it is possible to generate preferences that have an optimum with some bonds provided there is a small difference between annuity and bond returns, a small early withdrawal penalty for bonds and sufficiently actuarially unfair medical insurance pricing.

We turn now to a medical risk that occurs only in the second period. With medical insurance purchased at the start of period one, expected utility can be written as $(1 - m)U(c_1, R_A A + R_B B) + mU(c_1, R_A A + R_B B - M + \beta I)$. Thus, medical risk in the second period does not change the dominance of annuities over bonds whatever the pricing of medical insurance. In the absence of the arrival of information, there are two equivalent ways of organizing medical insurance. One is to purchase medical insurance at the start of period one (as with long-term care insurance). The other is to purchase an annuity with a plan to purchase medical insurance at the start of period two, if alive. With both formulations, a worsening of the pricing of medical insurance will generally alter first-period consumption and so total savings. In the second formulation, this translates directly into the demand for annuities. In addition, there would generally be a change in the amount of medical insurance purchased in the second period. In the first formulation, unlike the second, a change in the level of medical insurance would also change the level of annuity purchase even if preferences were such that first-period consumption did not change. For example, going from fair medical insurance to no medical insurance would increase the spending on annuities by the full amount previously spent on medical insurance if preferences were such that first-period consumption

did not change. That is, removing insurance that is effectively annuitized can increase the demand for a standard annuity.

Thus, we can conclude that the timing of the risk of medical expenses is key to understanding the interaction between the availability of medical insurance and the purchase of annuities. In the absence of strong assumptions it is thus impossible to sign the effect of liquidity needs on annuity demand. In one parameterization, Turra and Mitchell (2004) find that the optimal fraction of wealth annuitized remains large even when out-of-pocket (uninsured) medical expenditures are possible and are associated with truncated lifetimes. However, these simulations also show that these uninsured expenditures tend to reduce demand for annuities below 100 percent of savings.

We have assumed an absence of a relationship between medical expenses and life expectancy. If a medical expense in period 1 implies a lower survival probability, then that would strengthen the value of liquid bonds. The next subsection briefly considers a role of the arrival of information about life expectancy.¹²

Since we examine annuity demand in different settings, rather than a model of equilibrium, we do not explore the illiquidity of annuities (relative to bonds) that is present. Even if illiquidity of annuities were an explanation for lack of demand for illiquid annuities, Bernheim (1987b) has pointed out that illiquidity is not a complete explanation for the near absence of annuity markets. Bernheim proposes the creation of annuities that are subject to cancellation at any time. In terms of the notation above, bonds would have a return if held to maturity, R_B , and a return is withdrawn early, α_B , with annuities characterized by R_A and α_A . Even with an early withdrawal option, we might plausibly have $R_B < R_A$ and $\alpha_B > \alpha_A$. This could be the outcome since early withdrawal from an annuity has implications for the cost of providing the annuity, while this is less so of a bond (where withdrawal may just reflect available alternative investments).

C. *Inferior Returns to Annuities*

Another route to limited annuitization is if annuity pricing and the arrival of asymmetric information imply an advantage to delayed annuitization. For example, Milevsky and Young (2002)

¹²If the medical condition is observable to the provider, it may be that bundling insurance for the cash need with the annuity could improve pricing for both forms of insurance by eliminating adverse selection that would exist for either product individually, as has been proposed by Warshawsky, Spillman and Murtaugh (forthcoming).

consider a cost to annuitization associated with illiquidity that renders the return on annuities essentially inferior to the return on some bonds. In particular, they show that it may be optimal to wait to annuitize wealth if returns on investment in the future may exceed present returns. Similarly, if annuities purchased later in life are priced more favorably than those purchased earlier, deferral of annuitization may be optimal.¹³ If, however, there are utility gains to annuitizing at, say, age 65, and yet it is optimal to delay annuitization, then this implies that there are even larger utility gains to annuitizing later in life. Empirically, however, we do not observe households choosing to annuitize at later ages, suggesting that such an explanation cannot fully explain the near absence of demand for private market annuities.

Of course, if annuity returns, net of administrative costs, are inferior to those of conventional assets, annuity demand can go to zero. Mitchell et al (1999) show, however, that available pricing on annuities does not seem to be sufficient to render annuitization unattractive. Recent work by Dushi and Webb (2004), which builds upon previous work by Mitchell et al (1999) and Brown and Poterba (2001) shows that a combination of low annuity returns, within couple-risk sharing that substitutes for a formal annuity market, and very high levels of pre-existing annuities can render additionally fixed annuities unattractive at any age. However, even these results suggest that we should observe frequent annuitization by surviving spouses, which we do not.

III. Bequests

Throughout the paper we have maintained the assumption that there was no bequest motive. While a bequest motive reduces the demand for annuities, it does not eliminate it in general. To see this, consider a model with two periods and uncertain survival to the second period as the sole risk. Ignoring the role of inter vivos gifts, there can be a bequest at the end of the first period or of the second. We model the bequest motive as an additive term depending on the two levels of bequests that might happen at the end of periods one and two. In this case, we can express utility in terms of own consumption and the two possible bequest levels. With death at the end of period two, the

¹³Dennis (2004) also finds a gain to deferred annuitization for some price patterns.

bequest is received in period 3, involving additional interest. Using the dual formulation, we have:

$$\min_{c_1, A, B, c_2} : c_1 + A + B \quad (23)$$

$$\text{s.t.} : U(c_1, c_2) + V(R_B B, R_B(R_A A + R_B B - c_2)) \geq \bar{U}. \quad (24)$$

If the utility of bequest is the expectation of a concave value of the bequest, v , evaluated at the start of period three (whether given then or earlier), we have:

$$V(R_B B, R_B(R_A A + R_B B - c_2)) = qv(R_B R_B B) + (1 - q)v(R_B(R_A A + R_B B - c_2)), \quad (25)$$

where q is the probability that the individual dies before period 2. For there to be no annuitization, the value of an annuity must be less than the value of a bond, evaluated at zero annuities:

$$(1 - q)R_B R_A v'(R_B(R_B B - c_2)) < R_B R_B (qv'(R_B R_B B) + (1 - q)v'(R_B(R_B B - c_2))) \quad (26)$$

With $c_2 > 0$, this is violated for an actuarially fair annuity $(1 - q)R_A = R_B$, and for annuities close enough to fair.

If the annuity is actuarially fair, then

$$v'(R_B(R_A A + R_B B - c_2)) = qv'(R_B R_B B) + (1 - q)v'(R_B(R_B B - c_2)) \quad (27)$$

This implies $R_A A = c_2$, a point implicit in Yaari (1965). That is, annuitization equals second period consumption. Thus with this approach to bequests, the case for significant annuitization survives the presence of a bequest motive.¹⁴

IV. Simulations

The theory shows that while full annuitization is not guaranteed unless every bond is dominated by some annuity, an unrealistic assumption, it is also clear that partial annuitization is optimal in many settings. The theory itself is insufficient for answering practical policy questions that are

¹⁴The use of a "guarantee" with annuities is common. This often involves a minimum number of payments, resulting in a lower monthly annuity for a given purchase price. We note that this generates a random bequest compared with purchasing the same lower annuity without the guarantee and bequeathing (or giving before death) the reduced purchase price. In the case of fair annuities, the guarantee is a pure gamble and implausibly optimal. With selection issues, the demand depends on pricing and the guarantee may be part of addressing selection issues.

often centered on what fraction of savings is optimally annuitized in different market settings and with different preferences. Thus, a large simulation literature has arisen to address this. We briefly review this literature, and then extend it by conducting a “stress test” of the annuity valuation results by considering cases that are intentionally designed to create a more extreme mismatch between desired consumption and the annuity income stream than we would expect to see in practice. By showing that the optimal level of annuitization remains high even in what we consider to be extreme cases, we conclude that the virtual absence of voluntary annuitization cannot be easily explained the standard set of reasons that come from the life-cycle framework.

Friedman and Warshawsky (1990) and Mitchell, Poterba, Warshawsky and Brown (1999) are examples of the simulation literature that calculate utility gains from annuities under alternative assumptions. These papers, like many others, consider an individual a retired individual whose preferences can be represented by a utility function that exhibits constant relative risk aversion, exponential discounting, and intertemporal separability. They find that these consumers would accept a substantial reduction in wealth in exchange for access to actuarially fair annuity markets.¹⁵

Numerous subsequent papers have explored the gains from annuitization under a wider range of assumptions, including: uncertainty about future asset returns (Milevsky and Young (2002)), risk pooling within couples (Brown and Poterba (2003)), uninsured medical expenditures (Turra and Mitchell (2004), Sinclair and Smetters (2004)), actuarially unfair prices due to mortality heterogeneity (Brown (2003), Palmon and Spivak (2001)), and higher levels of pre-existing annuities (Dushi and Webb (2004)). In nearly every case, the annuity products considered in these papers are constrained in some way (i.e., annuity markets are incomplete), the most common assumption being that the annuity’s payment stream is fixed in real terms. Consistent with our theoretical findings, these papers find that there are substantial gains to some annuitization, but that full annuitization is not always optimal. Our theoretical results provide a framework in which such results can be interpreted: when full annuitization is sub-optimal, the settings were such that incomplete markets led to a mismatch between desired consumption trajectories and available annuity income paths. Even in these cases, however, the optimal level of annuitization was generally high.

¹⁵Another strand of the literature examines the “probability of a shortfall” when a consumer tries to self-insure rather than using formal annuity markets. Milevsky (1998) is a good example of this approach.

A. *Relaxing Additive Separability: A “Stress Test” of Annuity Valuation*

A key practical message from our theoretical results is that the welfare gains from annuitization depend on preferences mainly through the consumption trajectory. Thus, if one is seeking to reconcile the theory with the empirical smallness of voluntary annuity markets, one ought to be looking for situations in which there is a severe mismatch between the desired consumption trajectory and the income path provided by a limited set of annuities in an incomplete markets setting.

In this section, we extend the simulation literature by modeling preferences that lead to a severe mismatch with the goal of seeing whether such mismatches can plausibly explain the small size of the market. We do this by exploring a class of models that have the realistic property that preferences are not intertemporally additive and have the feature that different parameterizations, meant to reflect the heterogeneity of household financial positions at retirement, lead to very different time shapes of optimal consumption.

We consider a 65 year-old male with survival probabilities taken from the U.S. Social Security Administration for the year 1999, modified (to ease computation) so that death occurs for sure by age 100. His utility function is:

$$U = \sum_{t=65}^{100} (1 + \delta)^{-t} (c_t/s_t)^{1-\gamma} / (1 - \gamma) \quad (28)$$

When the parameter s_t is constant, this is the frequently modeled case of CRRA preferences. Following this literature, we set the real interest rate and discount rate (δ) equal to .03 and calculate the utility gains from annuitization. We set $\gamma = 2$ and find the consumption vector that solves the expenditure minimization problem numerically using standard optimization techniques.¹⁶

We begin by simply verifying that our model matches the results of the previous literature in the case in which s_t is fixed at a value of one, corresponding to the CRRA utility case. In this setting, which we denote as “Case 1” in Table 1, we find that it is optimal for an individual to fully annuitize all wealth. Moreover, with $r = \delta$, real annuities provide the optimal annuity stream.

A rough estimate of the magnitude of EVs can be obtained by observing the difference in trajectories between the unconstrained (circles and \times 's in figure 1, case 1) consumption plan and the constrained real annuity (triangles in figure 1) consumption plan. When optimal consumption is sharply decreasing, the constraints bind consumption away from the optimal path. In these

¹⁶Results for other levels of risk aversion are available in some of the papers listed above.

cases, the benefit of annuitization is relatively small because the sum of future consumption is relatively small and the gain is offset further by the constraints. When optimal unconstrained (zero annuitization) consumption is hump shaped and less steeply decreasing, there are greater benefits and the constraints impose less costs, so the net benefit to annuitization is greater.

What differentiates our more general set-up from prior work is that we can vary s_t in equation (1) so that the utility function exhibits an “internal habit,” which we can then adjust to create optimal consumption trajectories that differ markedly from the usual CRRA case. The intuition behind our utility function, taken from Diamond and Mirrlees (2000), is that it is not the level of present consumption, but rather the level relative to past consumption, that matters for utility. For example, life in a studio apartment is surely more tolerable for someone used to living in such circumstances than for someone who was forced by a negative income shock to abandon a four-bedroom house. In choosing how to allocate resources across periods, “habit consumers” trade off immediate gratification from consumption not only against a lifetime budget constraint, but also against the effects of consumption early in life on the standard-of-living later in life.

Following Diamond and Mirrlees (2000), we model the evolution of the habit as follows:

$$s_t = (s_{t-1} + \alpha c_{t-1}) / (1 + \alpha) \tag{29}$$

α is the parameter that governs the speed of adjustment of the habit level. When α is zero, the habit is constant and we are back in the additively separable case. As α approaches infinity, present habit approaches last period’s consumption. We select an intermediate speed of habit adjustment of the habit ($\alpha=1$). Away from zero, we find that changes to α make no substantive difference to our results, and therefore do not report results for a range of α values.¹⁷

When the habit evolves ($\alpha > 0$), present consumption increases the future standard-of-living. This increase in the standard-of-living reduces the level of utility and increases marginal utility in the future. With later consumption, there are fewer future periods that are adversely affected by an increased standard-of-living.

While some researchers have attempted to measure parameters of alternative habit formation models, largely involving “external habit” formation in calibration exercises, there is to our knowledge no empirical study that provides estimates of the initial standard s_{65} in our model. This is

¹⁷Additional results are available from the authors upon request.

not a major limitation, however, as our objective is not to carefully calibrate a realistic model of a representative lifetime optimizing consumer, but rather to create optimal consumption trajectories that create an intentionally more severe mismatch with the available annuity structure than we actually observe in available consumption data and to recognize the wide diversity in wealth at retirement relative to lifetime income. Doing so allows us to “stress test” the annuity valuations in order to see if one can plausibly explain the paucity of voluntary annuity purchases within a strictly rational model. We do so by considering initial standards of living that generate consumption paths that are both extremely friendly and extremely unfriendly to annuitization.

The first four columns of Table 1 indicate the case number and the relevant parameters that are varied across cases. Column (5) reports the fraction of wealth that is optimally placed in a constant real annuity instead of bonds. In column (6), we report the equivalent variation (“EV”) associated with this optimal amount of real annuitization. This is the increase in wealth required to hold utility constant while moving from having the optimal amount annuitized in a real annuity, to having all of wealth in bonds. Column (7) reports the gains from annuitization (again as an EV) for the case in which the individual is permitted to choose an optimal payout trajectory, i.e., they are no longer constrained to purchase a constant real annuity. For shorthand, we refer to this as the “complete markets” result, although strictly speaking, all that is necessary is that the subset of annuities that are desired by this individual be available.

Because an internal habit introduces a new cost to early consumption, one might expect that the fixed real annuity will be even more desirable with an internal habit ($\alpha > 0$) than without. However, the accuracy of this intuition hinges on the level of the initial habit s_{65} relative to resources at retirement. If the initial habit is small, then it is correct that the presence of a potentially increasing habit leads to deferred consumption and increased valuation of the annuity. However, if the initial habit is so large as to be unsustainable, so that the habit level must fall over time, then the presence of the habit will push optimal consumption earlier, because of the desire to smooth the ratio of consumption to habit across time. If the habit decreases with time, then smoothing likewise requires that consumption decrease with time. Hence a very high initial habit relative to resources provides a mechanism beyond heavy discounting whereby the deferred consumption required by a constant real annuity may be very burdensome.

To calibrate the model, and to build some intuition, we first consider an individual who has

reached retirement with resources that, when annuitized, are sufficient to exactly satisfy their habit level of consumption for the rest of their lifetime. In this scenario, labeled “Case 2” in Table 1, the individual enters retirement with a habit level s_{65} that is precisely equal to the annual annuity value of their wealth.¹⁸ As expected, the presence of a potentially increasing habit leads to deferred consumption, i.e., and upward sloping consumption path in the early years (see Figure 1, Case 2). When only real annuities are available, the optimal allocation is to fully annuitize, because doing so provides the individual with the highest possible return without imposing any binding liquidity constraints. While the utility gain is even higher than in our base case, the utility gains are higher still if the individual is able to purchase annuities in a complete market, and thus exactly match the desired consumption trajectory. In this latter case, the ability to choose an optimal annuity trajectory is equivalent in utility terms to nearly doubling of non-annuitized wealth.

Reaching retirement with precisely the resources required to maintain one’s past living standard is not the empirical norm, however. Indeed, given the wide range of retirement resources across the population, it would not make sense to consider a common level of pre-retirement consumption relative to retirement resource across the entire population. Hurd and Rohwedder (2004) provide new evidence on the distribution of the percentage change in spending from pre to post retirement. They find a mean consumption drop of 13 - 14 percent. At the 20th percentile, they find a 30 percent decline in consumption, whereas at the 95th percentile they find a 20 percent increase in consumption.¹⁹ We choose to examine even more extreme points in the consumption distribution by evaluating the case in which the habit level of consumption is 50 percent and 200 percent of the amount that can be sustained by one’s retirement wealth.

In Case 3 of Table 1 and Figure 1, the individual has sufficient wealth to purchase a real annuity stream that is double the initial habit level in retirement. As in cases 1 and 2, it is still optimal to fully annuitize all wealth. Doing so in a constant real annuity generates a utility gain that is equivalent to a 67 percent increase in non-annuitized wealth. The ability to match annuities to the desired upward sloping consumption trajectory is even more valuable.

More interesting for purposes of our “stress test” is to examine the case in which the individual only has half of the level of wealth that would be required to sustain their initial habit consumption

¹⁸Given our mortality assumptions, the annual annuity value of \$100 of starting wealth is \$8.59.

¹⁹Hurd and Rohwedder do not report points in the distribution below the 20th percentile.

level. As can be seen in Case 4, this leads the individual to want to sharply reduce consumption early in retirement, in order to rapidly bring down the level of habit to a more sustainable level. As such, the optimal fraction of wealth held in a real annuity declines to 90 percent, as the individual uses the 10 percent of non-annuitized wealth to supplement consumption in the first 10 years, and then consumes the annuity for the remainder of life after bring the habit down to a lower level.

Pushing the “stress test” even further, Case 5 adds a high discount rate to the low wealth-to-habit ratio. By holding r fixed at .03 and raising the discount rate to .10, the optimal consumption trajectory, as shown in Case 5, is even more heavily front-loaded, and thus the mismatch with the real annuity income stream is particularly severe. Even so, the individual would optimally annuitize three-quarters of their wealth. In results not reported, we have found that even when the individual’s habit is so high that retirement wealth can provide for an annuity that is only one-sixth of the initial habit level, the optimal fraction of wealth invested in a real annuity is still nearly two-thirds of initial retirement wealth.

These simulations indicate that it is extremely difficult to generate optimal consumption profiles such that the optimal fraction of wealth annuitized drops much below two-thirds of initial wealth at retirement without appealing to many additional factors. This suggests that the absence of annuitization outside of Social Security and defined benefit pensions cannot be easily explained within a rational life-cycle model, even with preferences that lead to a severe mismatch between desired consumption and available annuity paths.

V. Conclusions and Future Directions

With complete markets, the result of complete annuitization survives the relaxation of several standard, but restrictive assumptions. Utility need not satisfy the von Neumann-Morgenstern axioms and need not be additively separable. Further, annuities need not be actuarially fair, but only must offer positive net premia over conventional assets.

With incomplete markets, the full annuitization result can break down when there is a sufficient mismatch between the optimal consumption path and the income stream offered by the annuity market. In the much-studied case of a world where only individual mortality is uncertain, we find that there may be considerable individual heterogeneity in the value of annuitization. Heterogeneity

in annuity valuations is driven by heterogeneity in the willingness to substitute late consumption for early consumption. We find that even for preferences that stretch the bounds of plausible impatience, a large fraction of wealth is optimally placed in a constant real annuity.

In our simulations, we have retained the abstractions of no bequest motive, no risks other than longevity and no learning about health status or other liquidity concerns. Exploring the consequences of dropping these assumptions in the context of non-separable preferences and unfair annuity pricing would be an important generalization, but obtaining results will require strong assumptions on annuity returns, on the nature of bequest preferences and liquidity needs, and on the stochastic structure of bond returns.

It is sometimes argued that the lack of annuity purchase is evidence for a bequest motive. This raises the question of what sort of bequest motive would call for an absence of annuities. If there is no annuitization, then a bequest is random in both timing and size, measured as a present discounted value. Assuming one cares about the risk aversion of recipients, this may be dominated by giving the heirs a fixed sum at a fixed time and annuitizing the rest. More generally, partial annuitization can reduce the variation in the bequest.

The near absence of voluntary annuitization is puzzling in the face of theoretical results suggesting large benefits to annuitization. While incomplete annuity markets may render annuitization of a large fraction of wealth suboptimal, our simulation results show that this is not the case even in a habit based model that intentionally leads to a severe mismatch between desired consumption and the single payout trajectory provided by an incomplete annuity market. These results suggest that lack of annuity demand may arise from behavioral considerations and that mandatory annuitization may be welfare increasing. It also suggests the importance of behavioral modeling of annuity demand to understand the equilibrium offerings of annuity assets.

References

- Abel, Andrew**, “Asset pricing under habit formation and catching up with the Joneses,” *American Economic Review, Papers and Proceedings*, 1990, 80, 38–42.
- Auerbach, Alan J. and Laurence J. Kotlikoff**, “Life Insurance of the Elderly: Adequacy and

Determinants,” in “Work, health, and income among the elderly” Studies in Social Economics series, Washington, D.C.: Brookings Institution, 1987, pp. 229–67.

– **and** – , “The Adequacy of Life Insurance Purchases,” *Journal of Financial Intermediation*, 1991, 1 (3), 215–41.

Bernheim, B. Douglas, “Dissaving After Retirement: Testing the Pure Life Cycle Hypothesis,” in Z. Bodie, J. Shoven, and D. Wise, eds., *Issues in Pension Economics*, Chicago: NBER-University of Chicago Press, 1987a, pp. 237–282.

– , “The Economic Effects of Social Security: Toward a Reconciliation of Theory and Measurement,” *Journal of Public Economics*, 1987b, 33 (3), 273–304.

– , “How Strong are Bequest Motives? Evidence Based on Estimates of the Demand for Life Insurance and Annuities,” *Journal of Political Economy*, 1991, 99 (5), 899–927.

– , **Lorenzo Forni, Jagadeesh Gokhale, and Laurence J. Kotlikoff**, “The Mismatch Between Life Insurance Holdings and Financial Vulnerabilities: Evidence from the Health and Retirement Study,” *American Economic Review*, 2003, 93 (1), 354–365.

Brown, Jeffrey R., “Redistribution and Insurance: Mandatory Annuitization with Mortality Heterogeneity,” *Journal of Risk and Insurance*, 2003, 70 (1), 17–41.

– **and James M. Poterba**, “Redistribution and Insurance: Mandatory Annuitization with Mortality Heterogeneity,” *Journal of Risk and Insurance*, 2003, 70 (4), 527–553.

– **and Mark J. Warshawsky**, “Longevity-Insured Retirement Distributions from Pension Plans: Market and Regulatory Issues,” Working Paper 8064, NBER 2001.

Campbell, John, “Asset Pries, Consumption and the Business Cycle,” in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Amsterdam: North-Holland, 2002.

Campbell, John Y. and John H. Cochrane, “By force of habit: A consumption-based explanation of aggregate stock market behavior,” *Journal of Political Economy*, 1999, 107, 205 – 251.

Constantinides, George, “Habit Formation:A Resolution of the Equity Premium Puzzle,” *Journal of Political Economy*, 1990.

- Deaton, Angus**, *Understanding Consumption*, Oxford, 1991.
- Dennis, Michael**, “A Theoretical and Empirical Comparison of Deferred and Immediate Annuities,” 2004. Mimeo, UC Berkeley.
- Diamond, Peter A. and James A. Mirrlees**, “Adjusting One’s Standing of Living: Two Period Models,” in P. J. Hammond and G. D. Myles, eds., *Incentives, Organization, and Public Economics, papers in Honour of Sir James Mirrlees*, Oxford: Oxford University Press, 2000.
- Diamond, Peter and Daniel McFadden**, “Some Uses of the Expenditure Function in Public Economics,” *Journal of Public Economics*, 1974, 3, 3–21.
- Dusenberry, James**, *Income, Saving, and the Theory of Consumer Behavior*, Cambridge: Harvard University Press, 1949.
- Dushi, Irena and Anthony Webb**, “Household Annuity Decisions: Simulations and Empirical Analysis,” *Journal of Pension Economics and Finance*, 2004, 3 (2), 109–143.
- Friedman, Benjamin M. and Mark J. Warshawsky**, “The Cost of Annuities: Implications for Saving Behavior and Bequests,” *The Quarterly Journal of Economics*, February 1990, 105 (1), 135–154.
- Gomes, Francisco and Alexander Michaelides**, “Portfolio Choice With Internal Habit Formation: A Life-Cycle Model With Uninsurable Labor Income Risk,” *Review of Economic Dynamics*, 2003, 6 (4), 729–766.
- Huang, Ming**, “Liquidity shocks and equilibrium liquidity premia,” *Journal of Economic Theory*, 2003, 109 (1), 104–129.
- Hurd, Michael and Susann Rohwedder**, “The Retirement-Consumption Puzzle: Anticipated and Actual Declines in Spending at Retirement,” Working Paper 9586, NBER 2004.
- Milevsky, Moshe A.**, “Optimal Asset Allocation Towards the End of the Life Cycle: To Annuitize or Not to Annuitize?,” *Journal of Risk and Insurance*, 1998, 65 (3), 401–426.
- **and Virginia R. Young**, “Optimal Asset Allocation and the Real Option to Delay Annuity: It’s Not Now-or-Never,” 2002. Working Paper, York University.

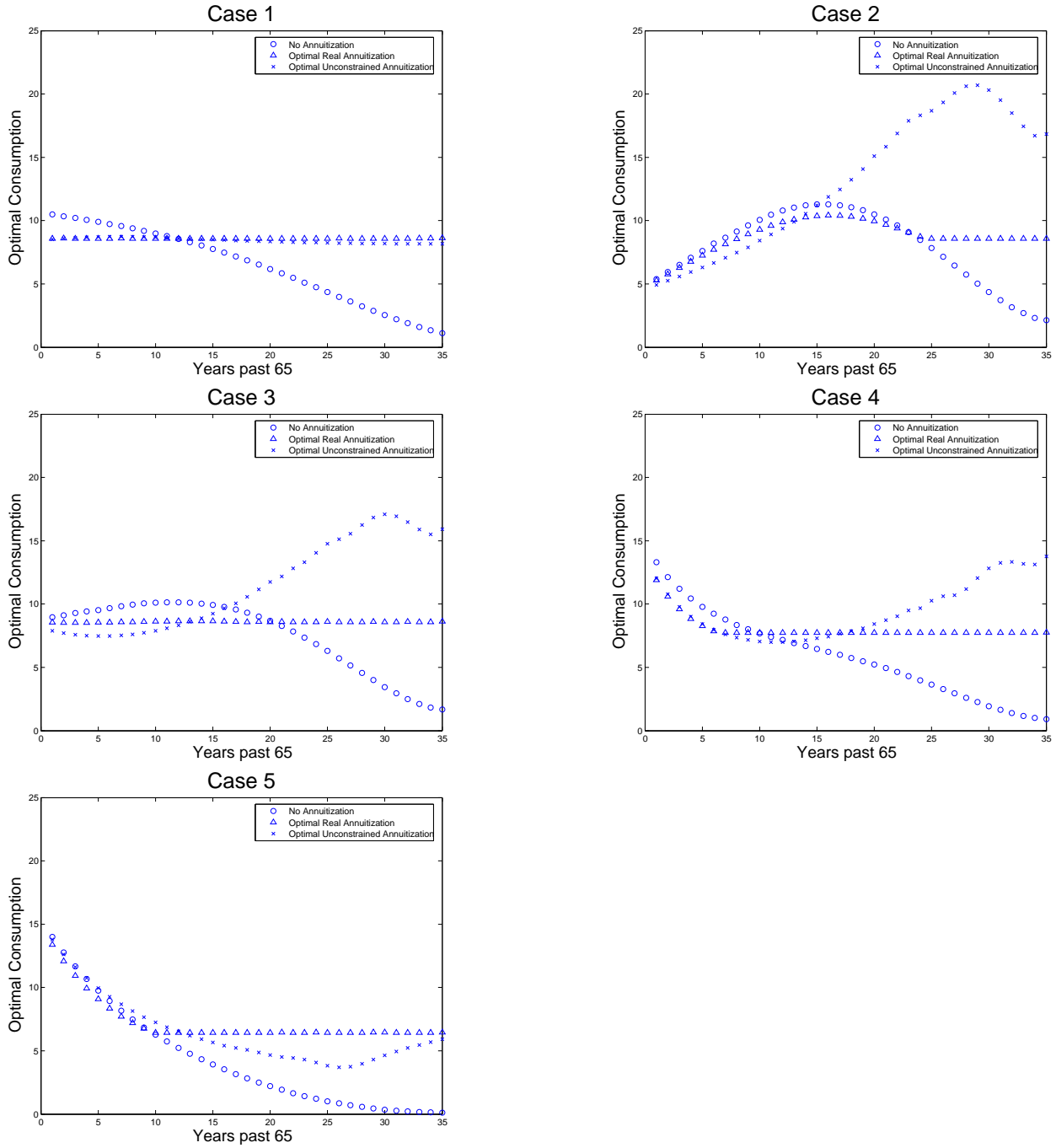
- Mitchell, Olivia, James M. Poterba, Mark Warshawsky, and Jeffrey R. Brown**, “New Evidence of the Money’s Worth of Individual Annuities,” *The American Economic Review*, December 1999, 89 (5), 1299–1318.
- Palmon, Oded and Avia Spivak**, “Information and the Market for Annuities,” Working Paper, Bank of Israel 2001.
- Sinclair, Sven H. and Kent A. Smetters**, “Health Shocks and the Demand for Annuities,” Technical Paper Series 2004-9, Congressional Budget Office 2004.
- Turra, Cassio M. and Olivia S. Mitchell**, “The Impact of Health Status and Out-of-Pocket Medical Expenditures on Annuity Valuation,” Working Paper, Pension Research Council 2004. WP 2004-2.
- Warshawsky, Mark J., Brenda Spillman, and Chris Murtaugh**, “Integrating the Life Annuity and Long-Term Care Insurance: Theory, Evidence, Practice and Policy,” in Olivia Mitchell, Zvi Bodie, Brett Hammond, and Steve Zeldes, eds., *Financial Innovations in Retirement Income*, forthcoming.
- Yaari, Menahem E.**, “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer,” *Review of Economic Studies*, April 1965, 32, 137.

Table 1. Simulated Utility Gains From Access To Annuitization

Case	Habit Adjustment Speed α	Ratio of initial habit s_{65} to real annuity	Discount rate	% of savings annuitized with real annuities	Welfare gain when optimal % of savings annuitized	Welfare gain when annuity markets complete
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	n/a	n/a	0.03	100	56	56
2	1	1	0.03	100	71	96
3	1	0.5	0.03	100	67	74
4	1	2	0.03	90	51	54
5	1	2	0.1	75	25	30

Notes: Risk aversion γ is set at 2 throughout. Column (3) compares the habit level (s_{65} in equation (28)) to the annual payout of an actuarially fair annuity with equal payouts in years 66 through 100. Welfare gains in columns (6) and (7) are calculated as the amount of initial wealth that a 65 year old male unable to annuitize any savings would have to be given to attain the same level of utility as if he were allowed to annuitize an optimal fraction of savings. Column (5) is the optimal fraction of savings placed in a real annuity, corresponding to the equivalent variation in column (6). Complete annuitization is optimal when the trajectory of payouts is unconstrained. Figure 1 plots optimal consumption trajectories with no annuities, real annuities and optimally designed annuities for each of the five sets of parameters.

Figure 1. Optimal Consumption Trajectories Under Different Annuity Availability 1



Notes: The five cases show correspond to the parameter values listed in Table 1. In each case, circles plot optimal consumption for a consumer who is unable to annuitize any wealth. Triangles plot optimal consumption when only a constant real annuity is available. \times 's plot optimal consumption when any consumption trajectory can be funded solely by annuities. The level of expenditures is the minimum to hold utility equal to the maximum utility when a constant real annuity is available and expenditures are equal to 100.