

# Optimal Annuitization with Stochastic Mortality and Correlated Medical Costs

Felix Reichling\*

Kent Smetters<sup>†</sup>

June 3, 2015

## Abstract

The conventional wisdom since Yaari (1965) is that households without a bequest motive should fully annuitize their investments. Numerous market frictions do not break this sharp result. We modify the Yaari framework by allowing a household's mortality risk itself to be stochastic due to health shocks. A lifetime annuity still helps to hedge longevity risk. But the annuity's remaining present value is correlated with medical costs, such as those for nursing home care, thereby reducing annuity demand, even without ad-hoc "liquidity constraints." We find that most households should not hold a positive level of annuities, and many should hold negative amounts.

*Keywords: Annuities, stochastic mortality, annuity puzzle, disability, health shocks.*

JEL Codes: D01, D14, H31

---

\*Felix Reichling: Congressional Budget Office. Email: Felix.Reichling@cbo.gov. The views expressed in this paper are the authors and should not be interpreted as those of the Congressional Budget Office.

<sup>†</sup>Kent Smetters: The Wharton School at The University of Pennsylvania and NBER. Email: Smetters@wharton.upenn.edu. Acknowledgments: The authors have benefited from discussions with Zvi Bodie, Michael Boskin, Jeffrey Brown, Daniel Gottlieb, Caroline Hoxby, Laurence Kotlikoff, John Laitner, Robert Merton, Rick Miller, Olivia Mitchell, Joshua Rauh, Dan Sacks, Tom Sargent, Florian Scheuer, Joseph Tomlinson, Stephen Zeldes, and seminar participants at the Boston University, BYU Computational Public Economics Conference, Congressional Budget Office, Stanford University, NBER, and the Q Group. We also received excellent comments from three anonymous referees, which significantly improved the exposition. We thank Leah Loversky for help with the HRS data. This research was supported by the Social Security Administration (SSA) through grant #5 RRC08098400-04-00 to the NBER as part of the SSA Retirement Research Consortium. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, CBO, SSA, or any other agency of the federal government. Parts of this paper was based on the unpublished working paper by Sinclair and Smetters (2004).

# 1 Introduction

The classic paper of Yaari (1965) demonstrated that the demand for annuities should be so strong that consumers without a bequest motive should invest all of their savings inside an annuity contract. Annuities are investment wrappers that should *statewise dominate* all non-annuitized investments because annuities produce a mortality credit—derived from the pooled participants who die and forfeit their assets—in addition to the return from the underlying principal. If an investor wants to invest in bonds then a *fixed*-return annuity invested in bonds will produce the bond yield plus a mortality credit. If an investor wants to invest in stocks then a *variable*-return annuity invested in stocks would produce the same realized yield plus a mortality credit.

Yaari's paper has received considerable attention because lifetime annuities, paying a fixed amount each age until death, are uncommon.<sup>1</sup> Indeed, the low annuitization of households is commonly referred to as “the annuity puzzle” (Modigliani, 1986; Ameriks et al., 2011). This puzzle is not just a theoretical curiosity. The mortality credit can be very large later in life, significantly increasing the return to saving. When an investor's preferences exhibit prudence, annuities also reduce the need for precautionary savings, improving consumption smoothing across the life cycle in the Yaari model.

As is well known, Yaari's model assumed costless and complete markets, and it ignored other types of longevity-risk sharing arrangements. In practice, annuity premiums incorporate sales charges and adjustments for adverse selection (Finkelstein and Poterba, 2004). People might face liquidity constraints after annuitization (Bodie, 2003; Davidoff, Brown and Diamond, 2005; Turra and Mitchell, 2008; Peijnenburg, Nijman and Werker 2013). Other sources of longevity pooling also exist, including Social Security and defined-benefit pensions (Townley and Boadway, 1988; Bernheim, 1991) and even marriage (Kotlikoff and Spivak, 1981).

Still, the careful analysis of Davidoff, Brown and Diamond demonstrates that many of these additional frictions do not undermine Yaari's full annuitization result. Brown et al. (2008, p. 304) conclude: “As a whole, however, the literature has failed to find a sufficiently general explanation of consumer aversion to annuities.” Indeed, as we show in this paper, Yaari's case for 100% annuitization of wealth is even more robust than commonly appreciated.

In the traditional Yaari model, an investor's survival is uncertain, and his or her probability of death naturally rises with age. But the mortality probability itself evolves deterministically over the life cycle for a given initial health status when young. The model does not allow for health shocks during the life cycle to suddenly change an investor's life expectancy.

In this paper, we largely adopt the Yaari framework but introduce health shocks that simultaneously affect longevity and increase uninsured health costs. Health shocks allow for the mortality probabilities themselves to be stochastic, a modification that is consistent with an investor's health status evolving over the life cycle with some randomness. Empirically, people accumulate precautionary savings to insure against random out-of-pocket health costs (Palumbo, 1999; French, 2005; De Nardi, French and Jones, 2010). It is natural to investigate how such shocks also affect decisions to annuitize. In our model, annuities continue to hedge longevity risk, as in the Yaari model and the large subsequent literature. But the presence of stochastic mortality probabilities also introduces a correlated risk. After a negative shock to health that reduces a household's life expectancy, the present value of the annuity stream falls. At the same time, a negative health shock produces potential losses, including lost wage income not replaced by disability insurance, out-of-pocket medical costs, and uninsured nursing care expenses, that

---

<sup>1</sup>Sales of fixed annuities in the United States totaled \$54.1 billion during the first three quarters of 2012, but only a fraction will be held for lifetime. See LIMRA (2012a).

may increase a household's marginal utility.<sup>2</sup> Since the value of non-annuitized wealth is not affected by one's health state, the optimal level of annuitization falls below 100%.

Some limited previous theoretical research has considered stochastic mortality probabilities (see, for example, the original work by Brugiavini, 1993). That work showed that agents will want to pool even this source of risk—known as “reclassification risk”—by purchasing the lifetime annuity early in the life cycle. However, that literature implicitly assumed that negative health shocks that decrease the value of the annuity are not correlated with any additional costs such as uninsured medical expenses.<sup>3</sup>

This paper shows that allowing for stochastic mortality *and* correlated longevity costs has a material impact on the optimal level of annuitization. We find that, relative to the conventional 100% annuitization result, more differentiated optimal life cycle annuitization patterns emerge from our model where: (i) most households do not annuitize any wealth; (ii) positive annuitization by non-wealthy households is largely concentrated in those households that can earn a large mortality credit relative to correlated risk (older households and sicker households); and (iii) positive annuitization is more likely in wealthy households where correlated costs are small relative to their assets. Full annuitization can be achieved if uninsured disability and medical costs are eliminated, assuming no informational asymmetries. Importantly, imperfect annuitization emerges even though households do not face any ad-hoc “liquidity constraints,” for example, because they can borrow against the present value of their annuity stream. Correlated costs also serve as an important gateway mechanism for other market frictions to reduce annuitization even more. One such friction is adverse selection. Its presence has no impact on the full annuitization result in the Yaari model (Davidoff, Brown and Diamond, 2005). The presence of selection, however, can reduce annuitization in our model, consistent with the fact that insurers actually do increase premiums for selection in practice.

We present simulation evidence using a multi-period life cycle model that is calibrated to the available data on household health and mortality risks, income loss, uninsured medical costs, and macro-level variables. Whenever we face data limitations for calibration, we err on the side of reducing the negative impact on annuity demand, including assuming that insurers have full information. Our simulation results demonstrate that annuitization is still typically much less than full, non-monotonic with age, and heavily influenced by the interactions of age-specific mortality and health uncertainty. We find that, depending on our calibration assumptions, between 64% and 76% of households should *not annuitize* any wealth, even with no transaction costs, bequest motives, ad-hoc liquidity constraints, or asymmetric information problems. In contrast, the Yaari model predicts 0%.

Our computed rates of annuitization are much lower than those reported in the previous literature that included health care expenditures (Davidoff, Brown and Diamond, 2005; Turra and Mitchell, 2008; Peijnenburg, Nijman and Werker 2013). Despite the seemingly commonality of health care shocks, the actual mechanism employed herein, however, is fundamentally different. In the previous literature, 100% annuitization emerges without liquidity constraints. With liquidity constraints, health care costs only served as a plausible motivation for lumpy expenditures during retirement. However, any type of

---

<sup>2</sup>Consistently, survey evidence shows that most people near retirement are concerned about the cost of long-term care. According to a Prudential Insurance Company survey, 74% of people between the age of 55 to 65 are concerned about needing long-term care in the future (Prudential, 2010) and that 63% of respondents are not confident about their ability to pay for LTC (Prudential, 2011).

<sup>3</sup>In Reichling and Smetters (2013), we also explore the theoretical role of the rate of time preference, where we show that imperfect annuitization can emerge even without correlated costs. The role of patience has been largely ignored in the annuity literature because, within the standard model with deterministic mortality probabilities, the discount rate on future utility only affects the level of saving and not the decision whether to actually annuitize that saving. With stochastic mortality, however, a standard annuity will fail to smooth consumption across those states that are actually most valued by investors. However, since we verify that this factor does not affect our simulation results, we have omitted this discussion for brevity.

lumpy expenditure, whether health care related or not, would produce incomplete annuitization with binding liquidity constraints. But as Davidoff, Brown and Diamond (2005) emphasize, liquidity constraints don't necessarily reduce annuitization that much, if any. The reason is that the most lumpy expenditures occur late in the lifecycle. While households at retirement will hold some "lumpy" assets (bonds) to finance lumpy expenditures that might happen soon after retirement, the larger relative return produced by the annuity is well suited for expenditures later in life. In the presence of transaction costs that reduce annuitization, distant lump expenditures can actually *increase* annuitization. Indeed, Davidoff, Brown and Diamond (2005) summarize (P. 1582): "In the absence of strong assumptions, it is thus impossible to sign the effect of liquidity needs on annuity demand." Moreover, as we explain in Section 2.2.2, liquidity constraints are difficult to justify in the Yaari framework where the mortality probabilities themselves are deterministic.

In sharp contrast, we don't impose liquidity constraints in our model. Health shocks are also absolutely central to annuity demand because negative health shocks reduce the remaining value of the annuity. Other forms of lumpy expenditures that are *uncorrelated* with health, in fact, have no impact on annuitization, a point that we verify in our simulation analysis reported later. Instead, a negative health shock to, for example, a new retiree can lead to an immediate loss of 35 percent of annuitized assets, even though our calibrated health state Markov process allows for the possibility of recovery back to better health states. As a result, even moderately risk averse agents may not want to annuitize any assets (a corner condition) in the presence of additional uninsured medical expenses, especially those households with few assets who are most at risk of entering high marginal utility states.<sup>4</sup>

In fact, under the most unconstrained version of our simulation model where households are allowed to short annuities (hold a negative position), many households will indeed choose to do so. As is well known, annuities and life insurance are opposite investments in one's longevity (Yaari, 1965; Bernheim, 1991). A short annuity position can be implemented by buying life insurance and reducing saving. Normally, the demand for life insurance is only positive in the presence of a bequest motive. In our model, younger households (and some older ones) short annuities even with *no* bequests motive.

Why do young households want to short annuities even without a bequest motive? Because they have little wealth, tend to be healthy, but still face a lot of uncertainty about their future health. If they *bought* an annuity, they could earn only a small mortality credit but would be accepting a large amount of correlated risk. We find that many younger savers should instead do the opposite: *pay* a small mortality credit by shorting annuities as a hedge against costly future negative realizations of health. Then, after a future realization of negative health information, this short position can be reversed by going long in an annuity that is cheaper than it would have been before the negative health shock.<sup>5</sup> The difference in the value of these short-long offsetting trades produces a net profit to the household that can then be used to supplement DI benefits (in case of the young) or to pay for any correlated uninsured health expenses (in case of retirees).

This paper does not intend to explain all of the stylized facts surrounding annuities, including the design of annuity contracts (Gottlieb, 2012) or whether households fully understand the annuity purchase decision (Brown et al., 2008; Beshears et al., 2012). Rather, our results are mainly intended to fundamentally recast the optimal baseline when the assumption of deterministic mortality probabilities

---

<sup>4</sup>Sections 5.2.2 and 5.2.4 explore additional potential differences between our model and some previous simulations. Previous simulations typically considered annuitization at retirement. Even when the total level of wealth at the point of retirement was calibrated to an empirically observed value, the underlying parameter values appear to imply larger values of post-retirement wealth than are empirically observed. That distinction matters since richer households are more likely to annuitize in both our model and in the liquidity constrained model.

<sup>5</sup>The mechanics are discussed in more detail in Subsection 2.2.2. In more recent times, life insurance policies can even be directly resold in the secondary market, a small but rapidly growing market known as life settlements.

in the Yaari framework is relaxed.

Still, it is interesting to know whether the standard expected utility model could rationally produce a low level of annuitization by using the gateway mechanism provided by the presence of stochastic mortality probabilities. This question is interesting not because we necessarily believe that agents are as highly rational as our model suggests. Rather, as in Milton Friedman’s classic billiard ball example, it is interesting to know whether households on average are maybe not making big mistakes after all, even if it is the result of some heuristics or even a bit of luck.<sup>6</sup>

To investigate this issue further, we introduce some additional real-world factors into our model in the Appendix, including asset management fees and bequest motives. Under these conditions, at least, nine out of ten households do not hold any annuities. We also examine the impact of counterparty risk and argue that most plausible additional model extensions would, if anything, reduce annuitization even more. In other words, although it is reasonable to argue that the standard annuity puzzle remains, one also cannot rule out the ability of a rational expectations model to produce a low positive demand for annuities. Moreover, the “true annuity puzzle” might actually be why we do not see more *negative* annuitization.

Interestingly, the presence of stochastic mortality probabilities is also consistent with another puzzle from the annuity literature. Both industry research and academic experimental evidence indicate that households typically view annuities as *increasing* their risk rather than reducing it. Brown et al. (2008) interpret this evidence as compatible with narrow framing. In our rational expectations model, however, the presence of stochastic mortality probabilities implies that annuities deliver a larger expected return (from the mortality credit) along with *more* risk (from the correlated costs). A greater level of risk aversion, therefore, *reduces* annuitization in our model. Our results are also consistent with Beshears et al. (2012) who find that people who are uncertain about big expenses during retirement value the greater spending flexibility of non-annuitized assets.

The rest of the paper is organized as follows. Section 2 develops a three-period model with deterministic survival probabilities and argues that Yaari’s 100% annuitization result is even stronger than previously understood. Section 3 then analyzes the role of stochastic survival probabilities in reducing annuity demand. Section 4 presents a multiple-period life cycle model and Section 5 presents simulation evidence that includes various frictions. Section 6 summarizes and concludes with a discussion of whether “medical annuities” could help restore more annuitization.

## 2 Three-Period Model

Consider an individual age  $j$  in health state  $h$  who can live at most three periods:  $j$ ,  $j + 1$ , and  $j + 2$ . The chance of surviving from age  $j$  to reach  $j + 1$  is denoted as  $s_j(h)$ , which is conditional on health state  $h$  at time  $j$ . State  $h$  is drawn from a countable set  $H$  with a cardinality exceeding 1. The Markov transitional probability between health states is denoted as  $P(h'|h)$ , where  $h \in H$  is the current state and  $h' \in H$  is the state in the next period.

An annuity contract with a single premium  $\pi_j$  at age  $j$  is available that pays 1 unit in each future period  $j + 1$  and  $j + 2$ , conditional on survival. We can think of each payment as a constant real amount, much like the annuity originally considered by Yaari and most of the subsequent literature.

In a competitive environment where insurers can pool idiosyncratic mortality risk without additional transaction costs, annuities are fairly priced. The premium paid at age  $j$  must equal the *actuarial present*

---

<sup>6</sup>For example, Scholz, Seshadri and Khitatrakun (2006) demonstrates that many households appear to be saving close to optimal levels despite the complexity associated with such a decision.

value of the payment of 1 received in periods  $j + 1$  and  $j + 2$ :

$$\begin{aligned}
\pi_j(h) &= \frac{s_j(h) \cdot 1}{(1+r)} + \frac{s_j(h) \cdot \sum_{h'} P(h'|h) s_{j+1}(h') \cdot 1}{(1+r)^2} \\
&= \frac{s_j(h) \cdot 1}{(1+r)} \cdot \left( 1 + \frac{\sum_{h'} P(h'|h) s_{j+1}(h') \cdot 1}{(1+r)} \right) \\
&= \frac{s_j(h)}{(1+r)} \cdot \left( 1 + \sum_{h'} P(h'|h) \pi_{j+1}(h') \right)
\end{aligned} \tag{1}$$

where  $h' \in H$  is the health state realized in period  $j + 1$ . Notice that the premium paid at age  $j$  is conditioned on the health status  $h$  at age  $j$ , which implies that insurers can observe the household's health status (we consider the impact of asymmetric information later). The term  $\sum_{h'} P(h'|h) s_{j+1}(h')$  on the right-hand side of equation (1) is equal to the expected chance of surviving to period  $j + 2$ , which recognizes that health status can change between ages  $j$  and  $j + 1$ . The algebraic manipulation shown in equation (1) then allows us to write the premium price recursively, so that at age  $j + 1$ :

$$\pi_{j+1}(h) = \frac{s_{j+1}(h)}{(1+r)}, \tag{2}$$

where we use the fact that  $\pi_{j+2}(h) = 0$  because  $j + 2$  is the maximum lifetime.

The realized (ex post) gross annuity rate of return, denoted as  $1 + \rho_j(h)$ , is derived similar to any investment: the dividend yield (1, in this case) plus the new price ( $\pi_{j+1}(h')$ ), all divided by the original price ( $\pi_j(h)$ ). The *net* return for a survivor to age  $j + 1$ , therefore, is:

$$\rho_j(h'|h) = \frac{1 + \pi_{j+1}(h')}{\pi_j(h)} - 1. \tag{3}$$

## 2.1 Deterministic Survival Probabilities (The Yaari Model)

In the Yaari model, mortality is uncertain. But the mortality *probabilities* themselves are deterministic, which can be viewed as a restriction on the stochastic survival probability process, as follows:

$$P(h'|h) = \begin{cases} 1, & h' = h \\ 0, & h' \neq h \end{cases} \tag{4}$$

But, survival probabilities are *not* restricted to be constant across age. For a person with health status  $h$  we can allow for standard life cycle “aging” effects:

$$s_{j+1}(h) < s_j(h) < 1$$

In other words, the likelihood of survival can decrease with age in a manner that is fully predictable by initial health status  $h$  and the current age alone. (The second inequality simply recognizes that some people die.) However, the probabilities themselves are not stochastic because  $h$  is fixed.

Inserting equation (4) into equation (1), the premium for a person of health status  $h$  at age  $j$  is:

$$\pi_j(h) = \frac{s_j(h)}{(1+r)} \cdot (1 + \pi_{j+1}(h)) \tag{5}$$

which implies:

$$\frac{(1+r)}{s_j(h)} = \frac{1 + \pi_{j+1}(h)}{\pi_j(h)}$$

The realized net rate of return to an annuity, therefore, is equal to

$$\begin{aligned} \rho_j(h) &= \frac{1 + \pi_{j+1}(h)}{\pi_j(h)} - 1 \\ &= \frac{(1+r)}{s_j(h)} - 1. \end{aligned} \tag{6}$$

Notice that the realized annuity return shown in equation (6) is identical to that of a single-period annuity—that is, it is independent of the survival probability at age  $j + 1$ . Intuitively, the survival probability at age  $j + 1$  is already known at age  $j$  and priced into the annuity premium  $\pi_j(h)$  paid at age  $j$ . It follows that a multiple-year annuity can be created with a sequence of single-period annuities, a well-known result in the literature.

We say that annuities *statewise dominate* bonds if  $\rho_j(h) > r$  for all values of  $h$ . In words, annuities always produce a better return than bonds for any state of the world. The following result implies that annuities should be held by *all people for all wealth* in the Yaari economy.

**Proposition 1.** *With deterministic survival probabilities and no bequest motive, fairly-priced annuities statewise dominate bonds for any initial health state at age  $j$ . (Proof is in Appendix A.)*

Statewise dominance is the strongest notion of stochastic ordering. Any person with preferences exhibiting positive marginal utility (including even very non-standard preferences that place weight on *ex post* realizations) prefers a statewise dominant security. Statewise dominance implies that annuities are also *first-order* dominant (hence, will be chosen by all *expected utility* maximizers) and *second-order* dominant (hence, will be chosen by all *risk-averse expected utility* maximizers).

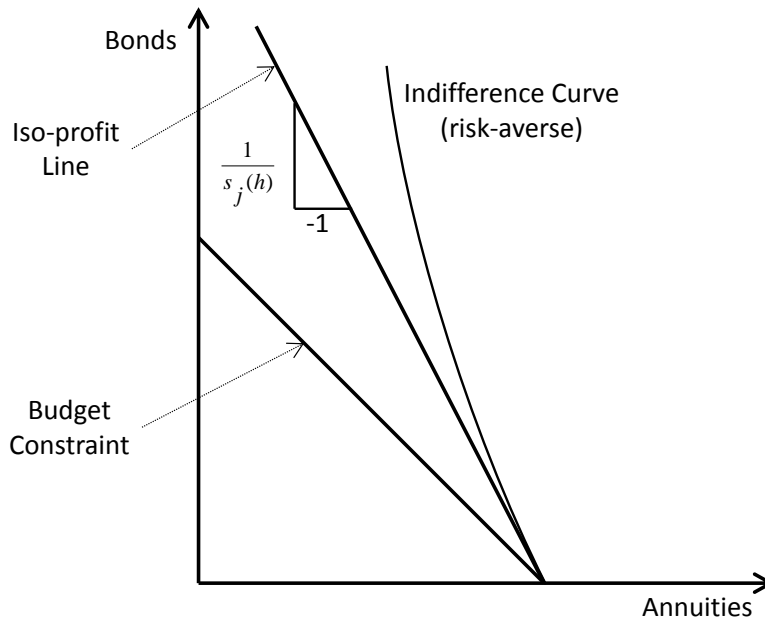
## 2.2 Robustness

It is well known that Yaari’s full annuitization result is robust to many market frictions (see, for example, Davidoff, Brown and Diamond, 2005). But the case for full annuitization is even stronger than commonly appreciated. Understanding the strength of the Yaari result allows us to understand the role that stochastic survival probabilities play in providing a gateway mechanism for many common market frictions to reduce annuitization. Toward that end, we present some novel graphical analysis that helps illuminate the robustness of annuities in the Yaari model.

Figure 1 gives some graphical insight into the statewise dominance in the Yaari model. Consider an investor at age  $j$  who is deciding between investing in bonds or buying an annuity with a competitive return that is conditional on her health  $h$  at age  $j$ . Her “Budget Constraint” between bonds and annuities is simply a straight line with slope of  $-1$ : she can either invest \$1 into bonds or \$1 into annuities.

The linear “Iso-profit Line” in Figure 1 shows the tradeoff between bonds and fairly priced annuities that would be offered by a competitive annuity market. The slope of the Iso-profit Line is steeper than the budget constraint and is equal to  $-\frac{1}{s_j(h)}$ . In words, it takes  $\frac{\$1}{s_j(h)} > \$1$  invested into bonds at age  $j$  to produce the same level of assets at age  $j + 1$  as \$1 invested into an annuity. Mathematically,  $\$1 \cdot (1 + \rho_j) = \frac{\$1}{s_j(h)} \cdot (1 + r)$ , as shown in equation (6).

Figure 1: Optimal Annuitization in the Yaari Model



The Iso-profit Line is also the Indifference Curve between bonds and annuities for a *risk-neutral* investor.<sup>7</sup> Specifically, a risk-neutral investor would be willing to give up \$1 in annuity investment if she could trade it for  $\frac{\$1}{s_j(h)} > \$1$  worth of bonds, because both investments would have the same value at age  $j + 1$ . Of course, the bond market would not allow for this trade, as indicated by the flatter budget constraint. The maximum Indifference Curve that can be achieved by a risk-neutral investor, therefore, must intersect the budget constraint at the corner point of full annuitization, as shown in Figure 1.

For completeness, Figure 1 also shows the “Indifference Curve (risk-averse)” for a risk-averse agent. Its slope must be at least as steep as the Iso-profit line, because a risk-averse agent would require *at least*  $\frac{\$1}{s_j(h)}$  worth of bonds to remain indifferent to a \$1 reduction in annuity protection. Risk-averse investors, therefore, also fully annuitize, as Yaari showed. Intuitively, a risk-averse investor values both the mortality credit and the enhanced consumption smoothing that the annuity provides.<sup>8</sup>

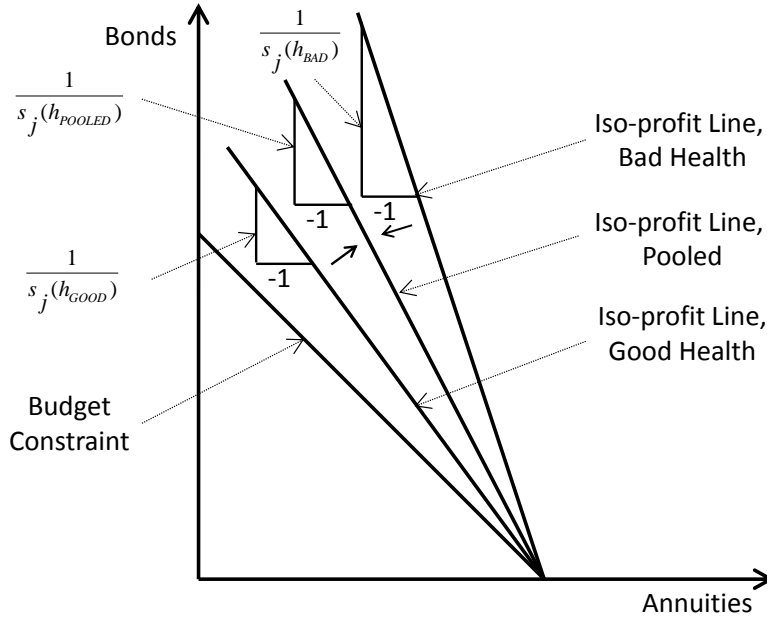
The “corner optimality” of the Yaari model is hard to break. Appendix B shows the robustness of the corner optimality to a host of market imperfections, including social security, insurance within the marriage, moral hazard, and uncertain income. Appendix B also explains why the common “liquidity constraint” critique of full annuitization should more accurately be thought of as a constraint on “asset rebalancing” and that such a constraint can’t bind in the Yaari economy. While transaction costs can break full annuitization, it either produces the corner of zero annuitization or the corner of full annuitization.

<sup>7</sup>Incidentally, it is also the Indifference Curve for a risk *loving* investor since the maximum payoff to bonds is actually *lower* than the guaranteed payoff to an annuity in the Yaari model, due to the statewise dominance of annuities.

<sup>8</sup>At this point, we are being a little informal; we have not formally defined risk aversion. Also, by focusing on the two-dimensional asset choice, Figure 1 ignores the saving decision itself. The potential of annuities to inter-temporally smooth consumption creates additional value for risk-averse agents in the Yaari model, whereas risk-neutral agents only value the extra mortality credit. These details are more formally treated in Section 3.2.2 within a special case of our model. For our purposes right now, it is sufficient that the Iso-profit Line is the weak lower bound for any risk-averse agent’s Indifference Curve, because we can demonstrate the robustness of the Yaari model using only the Iso-profit Lines.



Figure 2: Optimal Annuitization in the Yaari Model with Adverse Selection



### 2.2.1 Example: Adverse Selection

As an example, Figure 2 illustrates how the corner optimality is also robust to the presence of adverse selection. Suppose that health  $h$  at age  $j$  can take on two states: Bad health,  $h_B$ , and Good health,  $h_G$ , where, naturally, the probability of survival is lower for bad health:  $s_j(h_B) < s_j(h_G)$ . Without adverse selection, the insurer can separately identify people with Bad health and Good health. With adverse selection, the insurer cannot distinguish. As shown in the last subsection, the Iso-profit Lines represent the *lower bound* of an Indifference Curve of a risk-averse agent. It follows that we can omit the Indifference Curves in order to reduce clutter and can work directly with the Iso-profit Lines to demonstrate the robustness of the 100% annuity corner.

Let us first consider the case without adverse selection, where insurers can identify an annuitant's health type. The Iso-profit Line in Figure 2 for Bad health shows the tradeoff between bonds and fairly priced annuities that a competitive annuity provider would assign to people with Bad health. Similarly, the Iso-profit Line for Good health shows the tradeoff for people with Good health. Naturally, the Iso-profit Line for Bad health is steeper because people with Bad health face higher mortality risk and, therefore, earn a competitively higher return. In other words, to give up \$1's worth of annuities, a person with Bad health requires a larger amount of bonds than does a person with Good health.

Now suppose that annuity providers cannot distinguish between people with Bad and Good health, seemingly creating the potential for adverse selection. Instead, a single annuity is offered at terms representing the population-weighted average of both risk types, as indicated by the Pooled Iso-profit Line in Figure 2.<sup>9</sup> The effect of this pooling is that households with Bad health experience a loss in annuity return, indicated by a downward rotation in their Iso-profit Line. Household with Good health experience a gain in annuity return, indicated by an upward rotation in their Iso-profit Line. But notice that full annuitization for both types still occurs, despite the cross-subsidy, because each Iso-profit Line

<sup>9</sup>Mathematically, suppose that  $x\%$  of people had Bad health and  $(1-x)\%$  had Good health. Then, the  $s_j(h_{POOLED}) = x \cdot s_j(h_{GOOD}) + (1-x) \cdot s_j(h_{BAD})$ . The pooled Iso-profit Line is also a competitive equilibrium, provided that there is no other annuity provider that can better identify the individual risk types.

still intersects the budget constraint at the point of full annuitization. Intuitively, although adverse selection reduces the size of the mortality credit for some households, a smaller mortality credit is still better than no mortality credit in the Yaari model.

## 2.2.2 Example: “Liquidity Constraints”

The presence of *binding “liquidity constraints”* has been commonly cited as another friction that would undermine the case for full annuitization in a Yaari type model that is augmented with uninsured expense shocks. Intuitively, if a household annuitizes its wealth, then the wealth can no longer be used to buffer shocks that would increase its marginal utility, because the annuity income is received slowly over the life cycle. In contrast, the principle of short-term bonds should be more accessible.

As we now argue, however, the presence of binding liquidity constraints is challenging to reconcile with the assumption of *deterministic* mortality probabilities, as in the Yaari model. Before getting to the crux of the argument, it is important to be specific with terminology.

In particular, the “liquidity constraint” argument in the annuity literature is actually very different from the standard *borrowing constraint* assumption found in most literature, where people cannot borrow against their *future income*. There is a well-established microeconomics foundation about why it is hard for people to borrow against their future risky human capital.<sup>10</sup> Incidentally, a borrowing constraint of this sort does not undermine the case for full annuitization: Any existing savings (even if precautionary) should always be invested in a statewise dominant security.

Instead, the “liquidity constraint” argument, as used in the context of annuities, is imposing a very different requirement, namely a constraint on *asset rebalancing*. For incomplete annuitization to occur, households must be unable (or only at a high cost) to rebalance their *existing assets* from annuities into bonds. This constraint has nothing to do with future income and is different than a standard borrowing constraint. It is also very difficult to rationalize in the Yaari model. Consistently, Sheshinski (2007, p. 33) writes that “no apparent reason seems to justify these constraints.”

Indeed, simple annuity-bond rebalancing *would be* competitively provided *if* there were no reclassification risk to survival probabilities, as in the Yaari model.<sup>11</sup> A household could simply rebalance at age  $j + 1$  by pledging the 1 unit of conditional annuity income received at ages  $j + 1$  and  $j + 2$  to a life insurance contract, and then borrow the present value of the life insurance contract,  $\pi_{j+1}$ . This loan has been fully collateralized against mortality default risk, and so it would be offered by a competitive market. There is no role for subsequent hidden information to undermine this loan in the Yaari model: If annuity providers could have estimated the initial survival probabilities (that is, health state  $h$ ) necessary for underwriting the original annuity for a person at age  $j$ , then they also know the mortality probabilities at age  $j + 1$  with perfect certainty, because those probabilities change in a deterministic manner with age in the Yaari model. Even the subsequent transaction costs would be trivial, because those costs result mostly from medical underwriting, which would be unnecessary.

Empirically, rebalancing may not seem prevalent, but that may be the result of the small size of the *primary* annuities market. There actually *is* a direct secondary market for retirement annuities, and it is not clear whether the *available supply* of buyers is small relative to the small number of primary transactions.<sup>12</sup> Moreover, as just noted above, a person can reverse an annuity simply by purchasing

---

<sup>10</sup>Most of the literature has focused on the inability of the private sector to fully enforce two-sided contracts in the presence of hidden information. See, for example, Zhang (1997) and Clementi and Hopenhayn (2006).

<sup>11</sup>Even surrender fees that are intended to reduce rebalancing would inefficiently distort marginal utility in the presence of non-reclassification shocks and, therefore, could not survive competition.

<sup>12</sup>We could not find any aggregate industry information on the secondary annuity market. However, firms such as J.G. Wentworth actively advertise to purchase retirement annuities, as do other firms. Moreover, many life settlement firms,

life insurance. Because the life insurance policy has been fully collateralized by the original annuity, the present value of the life insurance policy's face can be easily borrowed against in the Yaari model. Empirically, the secondary market for life insurance continues to grow at a rapid pace, expanding the ability for such borrowing.<sup>13</sup>

Furthermore, even if the ability to rebalance still seems a bit of a stretch, it is important not to mix the underlying models. Problems with rebalancing could occur only in the presence of *stochastic* mortality probabilities that eliminate the perfect predictability of the previous health underwriting information found in the Yaari model. With stochastic mortality probabilities, medical underwriting would have to be repeated when the household wants to rebalance its annuity–bond portfolio. Of course, in practice, this would come at an additional cost to reduce adverse selection.

Nonetheless, in the theoretical derivations and simulation evidence presented below, we allow for costless asset rebalancing in the presence of stochastic mortality probabilities. Our purpose is to demonstrate the power of stochastic mortality probabilities themselves in reducing annuity demand without an additional rebalancing constraint—especially a constraint with unclear empirical support. We show that the falling value of the annuity itself following a negative health shock can play a major role in reducing the demand for annuities. Our results of imperfect annuitization would be even stronger if we included additional underwriting costs when annuity assets were rebalanced.

### 3 Stochastic Survival Probabilities

We now introduce stochastic survival probabilities by allowing  $P(h'|h) > 0$  when  $h' \neq h$ .

#### 3.1 Stochastic Rankings

The presence of stochastic survival probabilities can break the statewise dominance of annuities.

**Proposition 2.** *With stochastic survival probabilities ( $P(h'|h) > 0$ ), annuities do not generically statewise dominate bonds. (Proof is in Appendix A.)*

Intuitively, the annuity premium at age  $j$  is set competitively by insurers equal to the present value of the *expected* annuity payments received at ages  $j + 1$  and  $j + 2$ , conditional on the health state  $h$  at age  $j$ . But a sufficiently negative health *realization*  $h'$  at age  $j + 1$  reduces the expected payout at age  $j + 2$ , producing a capital depreciation at age  $j + 1$  that is larger than the mortality credit received. In effect, the annuity contract now has valuation risk (or principal risk) similar to a long-dated bond.

The fact that annuities do not statewise dominate bonds in the presence of stochastic survival probabilities, however, only means that annuities will not necessarily be optimal across a wide range of preferences with a positive marginal utility. Annuities could still dominate bonds for expected utility maximizers. Indeed, annuities will be strictly preferred by *risk-neutral* consumers who care only about the greater return from the mortality credit.

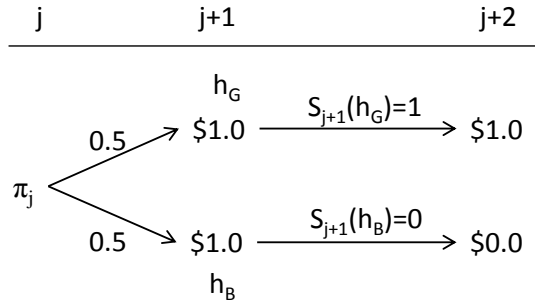
**Proposition 3.** *The expected return to a fairly-price annuity exceeds bonds if the chance of mortality is positive. (Proof is in Appendix A.)*

---

which normally buy life insurance policies, will also purchase annuities.

<sup>13</sup>See Life Insurance Settlement Association (2013). Moreover, pricing in the secondary market does not seem to contain large risk loads to compensate for systemic pricing mistakes. Most secondary transactions are medically underwritten, with many secondary life insurers now estimating life expectancy with a surprisingly high degree of accuracy. For example, Bauer and Russ (2012) used a database from a secondary life actuarial firm to construct a large panel of more than 50,000 individuals, finding small differences between originally estimated life expectancies and the actual death dates.

Figure 3: Annuity Payoff in Simple Example



However, annuities do not necessarily dominate bonds for the more restricted class of *risk-averse* expected utility maximizers.<sup>14</sup>

**Proposition 4.** *With stochastic survival probabilities, annuities do not generically second-order stochastically dominate (SOSD) bonds. (Proof is in Appendix A.)*

### 3.2 Examples

We now demonstrate Propositions 2 through 4 with a series of simple examples that build on each other.

#### 3.2.1 Failure of Statewise Dominance

Continuing with our three-period setting, consider an agent at age  $j$  with current health state  $h$ . Set the bond net return  $r$  to 0 to simplify the present value calculations. Also, assume that:

- At age  $j$ , a person with health state  $h$  will live from age  $j$  to  $j + 1$  with certainty ( $s_j(h) = 1.0$ ). Hence, the agent always collects the \$1 annuity payment at age  $j + 1$ .
- At age  $j + 1$ , an agent’s health status can take one of two states with equal probability:  $h_G$  (“Good”) and  $h_B$  (“Bad”). If the Good health state  $h_G$  is realized then the probability of surviving from age  $j + 1$  to age  $j + 2$  is one:  $s_{j+1}(h_G) = 1$ . Conversely, if the Bad health state  $h_B$  is realized then the probability of surviving from age  $j + 1$  to age  $j + 2$  is zero:  $s_{j+1}(h_B) = 0$ .

The payoffs for the annuity are summarized in Figure 3. By equation (1), the competitive annuity premium paid at age  $j$  is  $\pi_j(h) = \$1.0 + 0.5 \cdot \$1.0$ , or \$1.5. This amount is simply equal to the \$1 annuity payment received with certainty at age  $j + 1$  plus the expected value of the \$1 annuity payment received at age  $j + 2$ , which is paid 50% of the time to people who realize Good health at age  $j + 1$ .

Suppose that an agent at age  $j$ , therefore, is deciding between investing \$1.5 in the annuity or a bond.

- **Case 1 (“Good” health):** The household realizes the Good health state  $h_G$  at age  $j + 1$ . Then, by equation (3), the *realized* net return  $\rho_j(h_G)$  to the annuity is  $\frac{2}{1.5} - 1 > 0$ , thereby beating bonds which yield 0%. The annuity value of \$2 at age  $j + 1$  is equal to the \$1 annuity payment at  $j + 1$  plus the present value (at a zero discount rate) of the \$1 paid (with certainty) at age  $j + 2$ . Had this household instead invested \$1.5 at age  $j$  into bonds, it would have had only \$1.5.

<sup>14</sup>Specifically, gamble  $A$  second-order stochastically dominates gamble  $B$  if and only if  $E_{Au}(x) \geq E_{Bu}(x)$  for any nondecreasing, concave utility function  $u(x)$ .

- Case 2 (“Bad” health): The household realizes the Bad health state  $h_B$  at age  $j + 1$ . The realized net return to annuitization is equal to  $\frac{1}{1.5} - 1 < 0$ , under-performing bonds. Specifically, the annuity will pay only \$1 in total at age  $j + 1$ , the last year of life. Again, with bonds, the household would have had \$1.5.

It follows that annuities fail to statewise dominate bonds. Intuitively, the competitively priced annuity contract at age  $j$  was calculated based on *expected* survival outcomes at age  $j$ . Survival realizations below expectation must, therefore, leave some buyers worse off *ex post*.

### 3.2.2 Failure of Second-Order Dominance

The violation of statewise dominance, however, is only a small blemish for annuitization. It simply means that annuities will no longer be optimal across a wide range of preferences that, for example, place some weight on *ex post* realizations. Annuities could still be the dominant security for *risk-averse expected utility* agents whose preferences fully weigh risky gambles from an *ex ante* position—that is, at age  $j$ . Indeed, the presence of some *ex post* losers is the cost of *ex ante* risk reduction.

Let’s now consider the demand for annuities by risk-averse investors who care about smoothing consumption. Continuing with our example, we now explicitly introduce consumer preferences. We focus on the *standard expected utility* setting, where Yaari’s full annuitization result is standard and robust. Suppose that our agent at age  $j$  is endowed with \$1.5 and consumes only in ages  $j + 1$  and  $j + 2$ .<sup>15</sup> The agent has standard time-separable *conditional* expected utility preferences over consumption of

$$u(c_{j+1}|h_{j+1}) + \beta \cdot s_{j+1}(h_{j+1}) u(c_{j+2}|h_{j+1}), \quad (7)$$

where  $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$  takes the constant relative risk aversion (CRRA) form,  $\sigma$  is the level of risk aversion, and  $\beta$  is the weight placed on future utility. As will be evident below, our analysis holds for any risk-averse function, but the CRRA assumption allows us to report a few simple numerical examples.

The *unconditional* expected utility at age  $j$  is equal to

$$EU = \frac{1}{2} \cdot [u(c_{j+1}|h_G) + \beta \cdot u(c_{j+2}|h_G)] + \frac{1}{2} \cdot u(c_{j+1}|h_B), \quad (8)$$

where recall that  $s_{j+1}(h_{j+1} = h_G) = 1$  (i.e., Good health people live until age  $j + 2$ ) and  $s_{j+1}(h_{j+1} = h_B) = 0$  (i.e., Bad health people do not live to age  $j + 2$ ).

We consider two cases: without and with correlated costs:

**I) No Correlated Costs.** Continuing with our example, recall that an agent who buys an annuity at age  $j$  for \$1.50 and then realizes Good health at age  $j + 1$  will receive \$1 at age  $j + 1$  plus another \$1 at age  $j + 2$ . But, an annuitant who realizes a Bad health state receives \$1 only at age  $j + 1$ . In contrast, a bond investment simply returns the principle of \$1.5 at age  $j + 1$  because  $r = 0$ . Hence, with  $\beta = 1$ , the conditional consumption streams associated with these competing investment choices are:

- Bond: If Good health is realized at age  $j + 1$ , then  $c_{j+1} = 0.75$  and  $c_{j+2} = 0.75$ ; if Bad health is realized, then  $c_{j+1} = 1.5$ .<sup>16</sup>

<sup>15</sup>This timing is equivalent to a two-period model where the agent consumes in both periods and makes the investment decision prior to the update of survival probabilities at age  $j + 1$ .

<sup>16</sup>Specifically, if the Good health state is realized, then the agent lives periods  $j + 1$  and  $j + 2$  with certainty. Since  $\beta = 1$  and  $r = 0$  the agent simply splits \$1.5 between these two periods. If the Bad health state is realized then the agent lives only period  $j + 1$  with certainty, and so the agent simply consumes the \$1.5 fully in that period.

- Annuity: If Good health is realized at age  $j + 1$ , then  $c_{j+1} = 1.0$  and  $c_{j+2} = 1.0$ ; if Bad health is realized, then  $c_{j+1} = 1.0$ .<sup>17</sup>

Notice that the bond investment creates very non-smooth consumption choices across the two health states. In contrast, the annuity effectively shifts 0.5 units of consumption from the Bad health state to the Good state, thereby creating perfectly smooth consumption across states and time. Annuities, therefore, will be preferred by anyone with a reasonable felicity function  $u$  exhibiting risk aversion.

Moreover, notice that the agent will want to purchase the annuity at age  $j$  even though, by construction, the agent is guaranteed to survive from age  $j$  until  $j + 1$ ; the only uncertainty faced at age  $j$  is the *health state* the agent will realize at age  $j + 1$ . This result is consistent with the previous literature demonstrating that households will want to pool reclassification risk itself by contracting early in their lifetimes. See, for example, the original application of this result to annuities by Brugiavini (1993) as well as the excellent treatise by Sheshinski (2007, chapter 4).<sup>18</sup>

**II) With Correlated Costs.** Now consider the introduction of uninsured costs that are correlated with a decrease in survival probabilities. For example, a negative health shock can lead to a reduction in income (e.g., disability) and/or uninsured medical expenses (e.g., long-term care). Continuing with our example, suppose that a Bad health state is now associated with an additional loss of \$1 in the form of lower income or medical expenses. (There are no additional costs associated with Good health.) Now the consumption allocations for the bond and annuity investments are as follows:

- Bond: If Good health is realized at age  $j + 1$ , then  $c_{j+1} = 0.75$  and  $c_{j+2} = 0.75$ ; if Bad health is realized, then  $c_{j+1} = 0.5$ .
- Annuity: If Good health is realized at age  $j + 1$ , then  $c_{j+1} = 1.0$  and  $c_{j+2} = 1.0$ ; if Bad health is realized, then  $c_{j+1} = 0.0$ .

Under any felicity function satisfying the usual Inada condition ( $\frac{\partial u(c \rightarrow 0)}{\partial c} \rightarrow \infty$ ), the bond investment will now be chosen to avoid the possible zero consumption state that exists with the annuity. Of course, this example is intentionally extreme since correlated health costs fully absorb the annuity stream when health is Bad. With smaller correlated costs, partial annuitization would emerge. Simulation analysis is presented later using a more realistic calibration.

Notice that lower annuitization is not driven by any restriction on asset rebalancing. It just happens that once the Bad health state is revealed at age  $j + 1$ , the annuity produces no additional return at age  $j + 2$ , because the agent does not survive beyond  $j + 1$ . Appendix C discusses how this imperfect annuitization result can extend to variations in the design of the annuity contract.

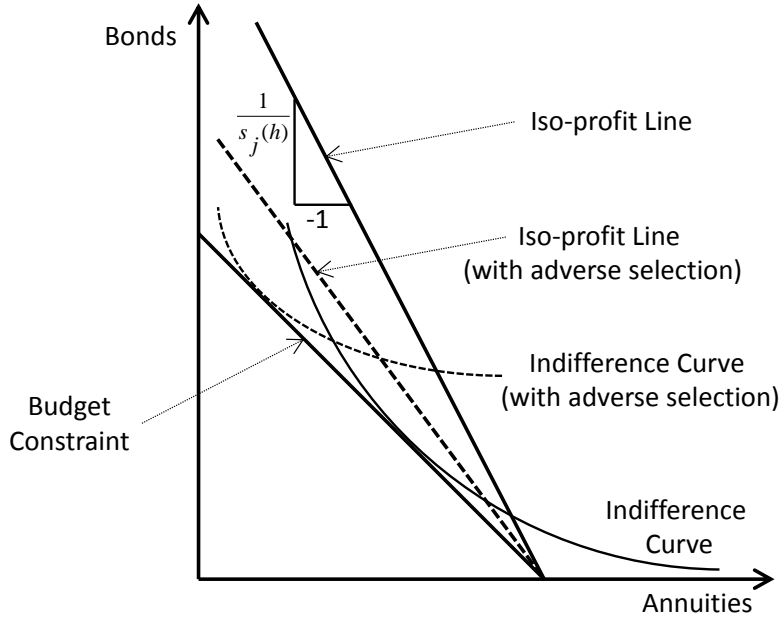
### 3.3 A Gateway Mechanism

The presence of stochastic mortality probabilities can remove the sharpness of 100% “corner optimality” found in the Yaari model. Figure 4 shows that the Indifference Curve between bonds and annuities

<sup>17</sup>Recall that the annuity in this example pays \$1 in each period that an agent survives.

<sup>18</sup>This literature, including the current paper, has focused on fairly priced contracts in the presence of household-level idiosyncratic shocks to longevity. Maurer et al. (2013) simulate a model with aggregate shocks to longevity common across households. Under a self-insure strategy, an insurer charges a load that reduces the probability that payments exhaust the insurer’s reserves, undermining fair pricing. Their results demonstrate the potential inefficiencies if government reserve regulation fails to properly weigh the insurer’s equity, reinsurance, and hedging contracts that would otherwise allow for a full transfer of risk from risk-averse households to risk-neutral insurers.

Figure 4: Optimal Annuitization with Health Shocks



can now take the more usual “convex-toward-the-origin” property. Recall that, in the Yaari model, the Iso-profit Line represents the lower bound of the Indifference Curve. However, in the presence of correlated uninsured health costs, an interior point for bonds might be selected. It follows that the Indifference Curve must contain at least one region that is flatter than the Iso-profit Line. Any reduction in the size of the mortality credit then further reduces the interior point demand for annuities. For example, the presence of adverse selection that reduces the mortality credit would rotate the Iso-profit Line downward, as shown with the dotted Iso-profit Line in Figure 4. The Indifference Curve must then rotate along the budget constraint, also shown in Figure 4, in order to ensure that at least one region of the Indifference Curve has a flatter slope than the corresponding Iso-profit Line. The net effect is a higher demand for bonds.

## 4 Multi-Period Model

We now present a multiple-period model followed by simulation evidence in Section 5.

### 4.1 Individuals

The economy is populated by overlapping generations of individuals who live up to  $J$  periods (years) with one-period survival probabilities  $s_j(h_j)$  at age  $j$  that are dependent on the realized health state  $h_j$ . Individuals are followed from the beginning of their working lives, through retirement, until death.

#### 4.1.1 Health Transition Probabilities and Conditional Survival Probabilities

The one-period survival probabilities  $s_j(h_j)$  are decreasing with age and health state. The health state  $h$  follows an  $M$ -state Markov process with an age-dependent transition matrix  $P_j(m, n)$ ;  $m, n = 1, \dots, M$ , where  $m$  is the current health state and  $n$  is the next health state. For our purposes, three states ( $M = 3$ )

Table 1: Fraction of Population in Nursing Homes: Data vs. Model

Age	Data	Model
65+	3.1%	3.1%
65-74	0.9%	0.9%
75-84	3.2%	3.2%
85-94	10.4%	10.4%

Source: U.S. Census Bureau (2011a).

Table 2: Distribution of the Duration of Nursing Home Stays: Data vs. Model

	Data	Model
Up to 1 year	44%	39%
1 to 3 years	30%	36%
3 to 5 years	14%	14%
5 years or more	12%	10%

Source: Centers for Disease Control and Prevention (2009).

*Explanation:* The Model values were *not* constructed to be equal to the empirical Data values. Instead, this table shows the average stay in nursing homes after adjusting the Robinson data to match the *fraction* of the population in nursing homes with Census data, shown in Table 1.

suffice: healthy ( $h_1$ ), impaired ( $h_2$ ), and very sick ( $h_3$ ). The impact of each health state on earnings and out-of-pocket costs depends on whether a person is retired or working.

**Retirees.** Retirees can transition between the healthy ( $h_1$ ) and impaired ( $h_2$ ) health states without any out-of-pocket costs. Retirees reaching health state  $h_3$  require nursing home care and face out-of-pocket expenses, but may receive, if qualified, transfers from Medicaid to cover those nursing home expenses. Survival probabilities and health transitions for retirees age 65 and older are based on the actuarial model of Robinson (1996).<sup>19</sup> His estimates are generally regarded as the industry standard for older individuals, as he carefully models the joint relationship between age, health status and longevity. We slightly adjusted Robinson’s health transition probabilities for states  $h_1$  and  $h_2$  so that the fraction of retirees in  $h_3$  match exactly the fraction of the population by age residing in nursing homes (see Table 1) as reported by the U.S. Census Bureau (2011a). Table 2 shows that these adjustments produce a plausible distribution of the duration of time spent in a nursing home conditional on admission. Robinson’s data

<sup>19</sup>Robinson’s model used eight health transition states, which we converted into three in order to increase the size of each state bucket for mapping to survival probabilities. The use of three states also serves the key economic determinants of our model. Our healthy state  $h_1$  corresponds to Robinson’s state 1 (no impairments),  $h_2$  corresponds to his states 2–4 (2: instrumental activities of daily living (IADL) impairments only; 3: 1 activity of daily living (ADL) impaired; 4: 2 ADLs impaired), and  $h_3$  corresponds to his states 5–8 (3+ ADLs impaired or some ADLs impaired and cognitive impairments), which he notes is consistent with additional medical expenditures. Future work could consider additional sick states beyond long-term care, provided that these states could be mapped to a specific set of chronic conditions and their associated uninsured costs. Currently, such data is not readily available.



Table 3: Disability Rates: Data vs. Model

Age	Data	Model
21-29	0.7%	0.7%
30-39	1.8%	1.8%
40-49	3.9%	3.8%
50-59	7.9%	7.9%
60-64	12.1%	12.1%

Sources: Social Security Administration (2012b, 2013b) and authors' calculations.

Table 4: Disability Transition Rates by Cause: Data vs. Model

	Data	Model
Recovery	7%	7%
Retirement	50%	44%
Death	40%	49%
Other	3%	0%

Sources: Congressional Budget Office (2012b), Social Security Administration (2011), and authors' calculations.

does not sample ages greater than 100. For ages greater than 100, we smoothly extend the Robinson data so that the average survival probabilities across health groups equal the aggregate survival probability from the 2009 Period Life Table (Table 4.C6 in Social Security Administration, 2013a).

**Workers.** While most annuity demand studies, including this one, focus on retirees, we also investigate the demand for annuities by workers for two reasons. First, as we show below, the wealth level is an important determinant of annuitization in our economy, and one standard metric to ensure a plausible amount of wealth is the economy's capital-output ratio. Second, we show that young workers might want to *short* annuities, which we allow as part of our robustness checks.

Robinson's data, however, does not sample younger workers. To construct survival probabilities for workers below age 65, we use data from the U.S. Social Security Administration. For the very sick ( $h_3$ ) workers, we use HP-filter smoothed survival probabilities for disabled workers estimated by Social Security actuaries (Table 11 of Social Security Administration, 2011). Unlike Robinson, Social Security data does not provide other survival probabilities conditional on health statuses, in particular, corresponding to  $h_1$  (healthy) and  $h_2$  (impaired). But the SSA does estimate population average survival probabilities across all health states (see, 2009 Period Life Table, Table 4.C6 in Social Security Administration, 2013a). Conditional survival probabilities can then be imputed using the survival probabilities for very sick ( $h_3$ ) workers, survival probabilities for the entire population, and information about the Social Security disability insurance eligible rates among the impaired and very sick population.

Health transitions for workers are initially based on Robinson's model but adjusted as follows. Naturally, most workers are healthy (state  $h_1$ ) and able to work. These workers, therefore, do not receive disability benefits or face *uninsured* institutional health care costs. Workers who transition to health

state  $h_2$  do not work and apply for disability benefits for one period. If a worker’s application is approved, the worker transitions from state  $h_2$  to state  $h_3$  and receives DI benefits in subsequent periods; applicants who are neither accepted nor appeal their rejection transition back to  $h_1$  the following period since they are presumed to have gotten better. While these transitions closely follow the U.S. disability law, they conservatively assume that the Social Security DI program does not make any errors by denying applications of people who cannot subsequently work.<sup>20</sup> They also conservatively assume that applicants who return to  $h_1$  do not suffer any additional productivity loss from being out of the labor force while they applied for benefits.<sup>21</sup> Transitions from  $h_2$  (applying for DI benefits) to  $h_3$  (receiving benefits) are based on acceptance probabilities of DI applications while the likelihood of remaining in  $h_2$  is based on the fraction of applicants who appeal the rejection of their applications.<sup>22</sup> Given this construction, transitions from  $h_1$  to  $h_3$  are set to zero so that workers cannot skip the application process; similarly, transitions from  $h_3$  to  $h_2$  are set to zero consistent with the assumption that a person only stops receiving DI benefits if they are healthy enough to work (rather than pushed into limbo).<sup>23</sup> Transitions from  $h_3$  to  $h_1$  are adjusted so that the aggregate DI recovery rate in the model matches the data reported by Congressional Budget Office (2012*b*) and Social Security Administration (2011) and shown in Table 4. Accordingly, some workers stay in  $h_3$  until they either transition into retirement at age 65 or die.<sup>24</sup> Finally, transitions from  $h_1$  to  $h_2$  are adjusted so that the model matches data on disability by age group as reported by Social Security Administration (2012*b*, 2013*b*).

This calibration matches the available data well. Table 3 reports the share of the working-age population that is disabled (in  $h_3$ ) in our model versus the data. Notice that our model disability rates almost exactly match the data. Table 4 then reports the transition out of disability by cause (recovery, retirement, death). Our model’s “recovery” rate is identical to the data. In comparison to the data, however, our model has slightly larger transitions into death relative retirement.<sup>25</sup> Our calibration also lines up well against available aggregate survival data shown in Figure 5.

#### 4.1.2 Investment Choices

**Bonds.** Households can invest in a non-contingent bond that pays a net return equal to  $r$ , which is equal to the marginal product of capital, as defined later. Bonds, therefore, constitute a safe investment in our model because the aggregate capital stock is deterministic.

**Annuities.** Households can also invest in an annuity that pays \$1 per unit contingent on survival. Figure 6 shows the realized single-period net annuity return  $\rho$  as a function of age and health-state

<sup>20</sup>In practice, the U.S. DI program has a high level of false rejections (Low and Pistaferri, 2010). A majority of claims are initially rejected; and only about 35 to 40 percent of all claims (new and old) are approved (Ohlemacher, 2013).

<sup>21</sup>Our mean waiting period is almost identical to the careful estimation in Autor et al. (2015). However, they also estimate a significant reduction in subsequent labor earnings from being out of the labor force, which we don’t capture.

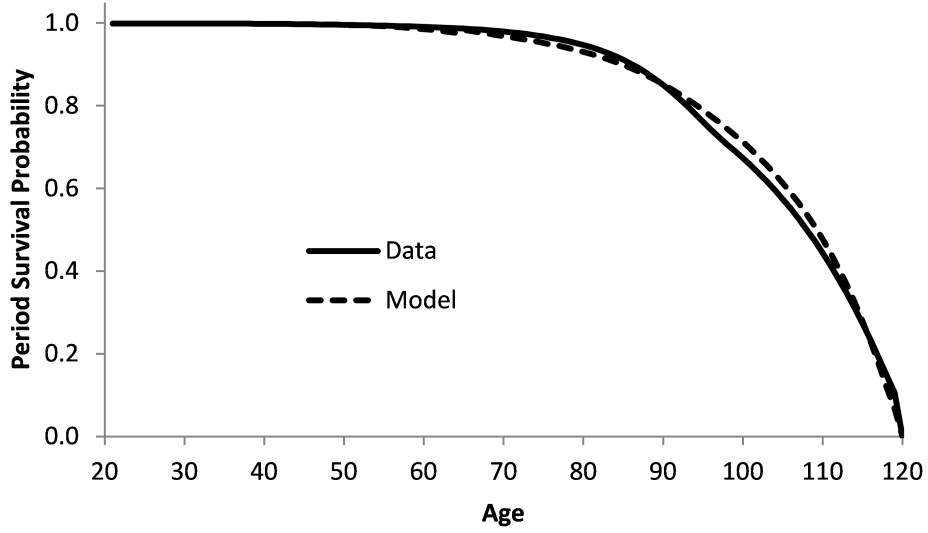
<sup>22</sup>See Congressional Budget Office (2012*b*) for a graphical representation of the DI application process and associated probabilities.

<sup>23</sup>Our numerical results are not highly sensitive to modest deviations from these assumptions. If we allow for an immediate transition from  $h_1$  to  $h_3$ , the calculated annuitization rate increases; if we allow for a positive transition from  $h_3$  to  $h_2$  then annuitization decrease. But, neither feature is consistent with the law. We also tested an alternative calibration of disability benefits where a fraction of people in  $h_2$  received benefits based on a lottery, but that made very little difference.

<sup>24</sup>To ensure that the fraction of the population of 65-74 year olds in nursing homes ( $h_3$  for retirees) match the data, we then transition all, but a small number of people from  $h_3$  to  $h_1$  and  $h_2$  at age 64, right before they enter retirement. In other words, consistent with the empirical data, most disabled people entering retirement do not immediately transition to a nursing home.

<sup>25</sup>We investigated ad-hoc adjustments on the numerical results we reported and found very small effects.

Figure 5: Aggregate Survival Probabilities: Data vs. Model



transition.<sup>26</sup> This return includes two components: (a) for a *given* health state, the annuity earns the standard *mortality credit* that increases over the lifecycle, and (b) after a *change* in health state, the annuity incurs a repricing: either a depreciation if health worsens or an appreciation if health improves. Notice, for example, that movements from healthy state  $h_1$  to the worsening health states  $h_2$  or  $h_3$  lead to large depreciations and often produce negative rates of return. Health movements in the opposite direction can lead to appreciation, although the probability of those shifts is less likely later in life.

### 4.1.3 Income and Expenses

An individual's income  $X_j$  at the beginning of age  $j$  is equal to

$$X_j = \varepsilon_j \eta_j I(h = h_1) w(1 - T) + B_j - L_j + Tr_j, \quad (9)$$

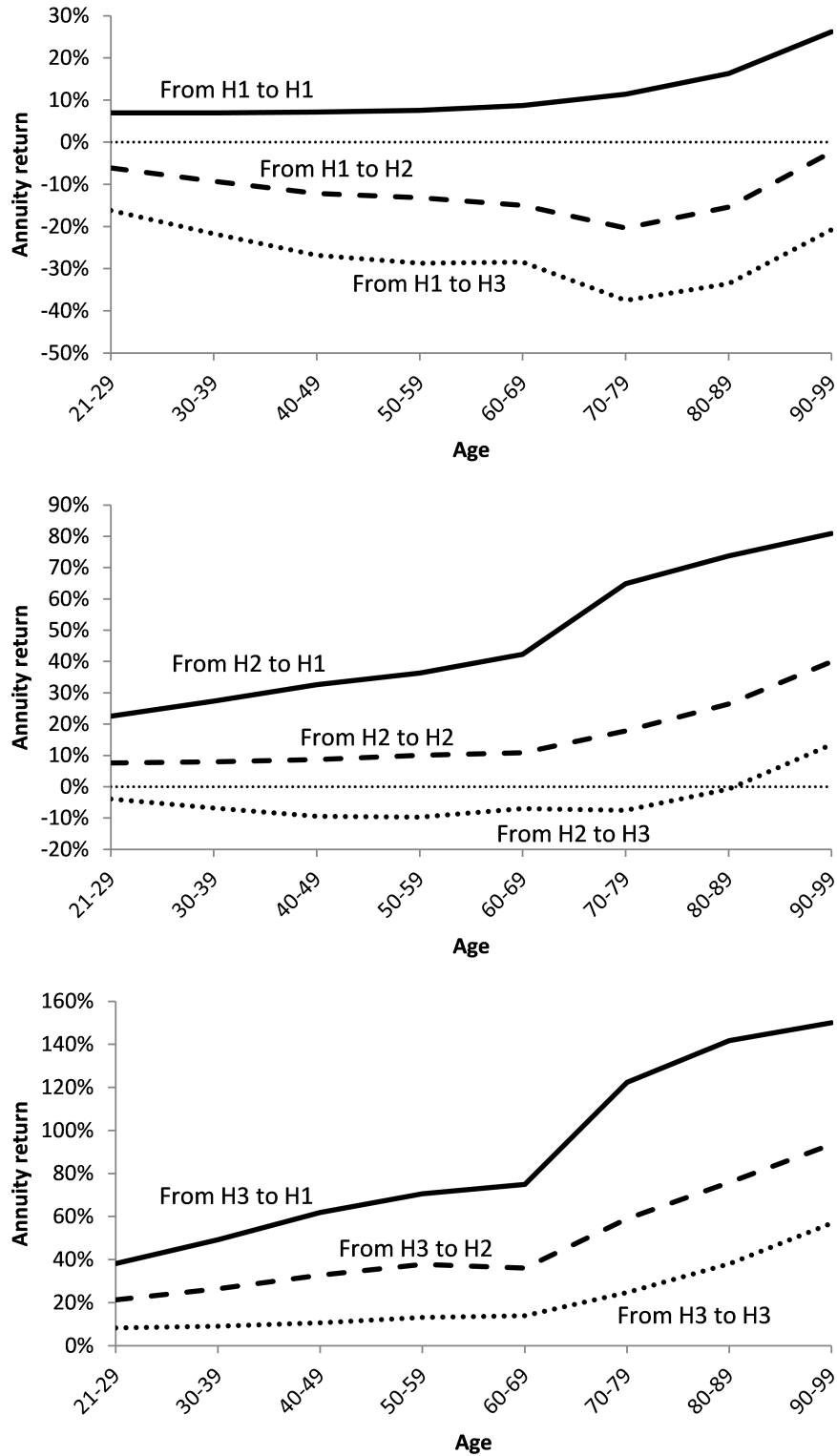
which includes after-tax wage income, bequests, uninsured long-term care costs, and government transfers. We now consider each factor in order.

**Wages and Disability.** Wages are a product of four factors:

- a predictable age-related productivity  $\varepsilon_j$  that is equal to the average productivity of a worker of age  $j$  (zero for ages  $j \geq 65$ , denoting retirement);
- an age-dependent individual random productivity  $\eta_j$  modeled as a Markov process with a transition matrix  $Q_j(k, l)$ ;  $k, l = 1, \dots, \Psi$ , where  $\Psi$  represents the highest productivity attainable in the economy;

<sup>26</sup>A shortcoming of our piece-wise connection of the transition data between workers (below age 65) using Social Security data and retirees (above age 65) using Robinson data is that it creates some kinks in the transition series  $H2 \rightarrow H1$  and  $H3 \rightarrow H1$  at age 65, as shown in Figure 6. We felt it was better to explicitly model and report these results rather than employ some ad-hoc smoothing. These kinks, however, have no impact on our first set of results reported below that focuses on annuitization at age 65 and throughout retirement.

Figure 6: Annuity Returns by Age and Health State Transition



*Explanation:* The panels show the annuity return  $\rho$  by age for different health-state transitions, calculated using equation (3). The top panel shows the annuity returns for a healthy ( $h_1$ ) person, the middle panel shows the annuity returns for an impaired person ( $h_2$ ), and the bottom panel shows the annuity returns for a very sick person ( $h_3$ ).

- an indicator of the health status  $I$ ; and
- the general-equilibrium market wage rate per unit of labor  $w$ .

The processes for  $\varepsilon_j$  and  $\eta_j$  are taken from Nishiyama and Smetters (2005), which allows for eight different earnings groups. The indicator function  $I(h = h_1) = 1$  if the person is healthy and able to work; otherwise,  $I(h \neq h_1) = 0$ , if the person is disabled ( $h_2$ ) or very sick ( $h_3$ ) and cannot work. The general-equilibrium wage  $w$  is produced using the technology described below.

**Bequests.** The variable  $B_j$  is the amount of bequests, positive in value for a bequest that is received and negative in value for a bequest that is given. Bequests of bond holdings are given at death; they are received earlier in life, typically by dividing up aggregate bequests evenly throughout the measure of the surviving population. We consider alternative bequest distributions as part of sensitivity analysis.

**Uninsured Medical Loss.** The variable  $L_j$  is the financial loss in the sick state  $h_3$ . During working years, its value is zero ( $L_j = 0$ ) under the assumption that all workers are privately insured. After retirement, the value of  $L_j$  is set to the value of nursing home costs of roughly \$83,000 per year, based on estimates by Genworth Financial (2012) and MetLife (2012), or equal to about 1.2 times the economy-wide average wage in the model economy. Medicaid pays for some of these costs for households that qualify (see below). This calibration conservatively assumes that at most only one of the household's retirees will use long-term care.

To be sure, there is a limited market for long-term care insurance. Historically, however, these policies have been expensive and offered only limited coverage (Brown and Finkelstein, 2011). Medicaid's provision of long-term care also crowds out demand for coverage by less affluent households (Brown and Finkelstein, 2008). Moreover, a qualifying insurable event is more subjective for long-term care than for life insurance or annuities, further complicating the purchase decision (Baldwin, 2013; Siegel Bernard, 2013).<sup>27</sup> Accordingly, only one in ten U.S. households have long-term care policies (Lockwood, 2013). This ratio is likely to further decrease as many of largest long-term care insurers—including Genworth, CNA Financial, Manulife, Metlife and Prudential—have recently stopped offering new coverage. The largest remaining provider, New York Life, has requested approval from state insurance regulators to substantially increase premiums (Lieber, 2010).

Besides long-term care, there are additional post-retirement medical expenses that are not reimbursed by Medicare, which covers only about 62 percent of health care services for eligible beneficiaries (Employee Benefit Research Institute, 2013). In its latest annual survey, Fidelity Benefits Consulting (2013) estimates that a typical couple retiring in 2013 will face almost \$220,000 in health care expenses that are not reimbursed by Medicare, although Employee Benefit Research Institute (2013) places the value at \$261,000. These costs do *not* include long-term care, dental, vision or over-the-counter medications. To be sure, some of these expenses can be insured through private Medigap policies. However, Medigap premiums are typically marked up 30% above costs, and only about 26.5% of retirees have such policies according to Starc (2012, Table 5). Moreover, even those retirees with Medigap policies still face large out-of-pocket expenses. For example, even with the most comprehensive Medigap coverage allowed by law, a single retiree in poor health should expect to pay \$10,000 in health care costs per year according to the Centers for Medicare and Medicaid Services (2013), not including long-term care, dental, vision, and over-the-counter medications. Retired couples face larger expenses.

---

<sup>27</sup>We thank Caroline Hoxby for this point.

Conservatively, we ignore these additional costs unrelated to long-term care in the baseline calibration of our simulation model. The technical reason is that we do not have a precise mapping between our health state Markov transition matrix and health costs unrelated to long-term care. However, we will consider different parametrizations of the financial loss  $L_j$  as part of our sensitivity analysis.

**Government Transfers and Taxes.** Public insurance at least partially offsets some of the disability losses suffered by workers and the uninsured medical losses suffered by retirees.

*Workers* receive a Social Security Disability Insurance (DI) payment before retirement, calculated using the legal and progressive “bend point” formula in the United States. We adjust those benefits slightly downward to match the fraction of DI recipients who are in poverty, based on data reported by U.S. Census Bureau (2012*b*). That downward adjustment decreases the generosity of DI benefits and thereby aligns the model generated benefits profile by age group more closely with that observed in the data.<sup>28</sup> Finally, that adjustment also ensures that aggregate DI benefits (as a fraction of GDP) are close to what is observed in the data. Workers are assumed to have full health insurance and face no out-of-pocket medical costs. *The only correlated costs faced by workers from a negative health care shock in our model, therefore, are from the portion of income that is not replaced by DI.*

*Retirees* receive Social Security benefits according to the “bend point” calculations contained in the law that allows for redistribution, in exchange for making proportional payroll contributions up until the payroll tax ceiling.<sup>29</sup> We slightly downward adjust Social Security benefits to better match the distribution of monthly payments as reported in Table 5.J6 of Social Security Administration (2012*a*).<sup>30</sup> Aggregate Social Security benefits are equal to 4.1% of GDP in our model, compared to 3.8% of GDP in 2011.<sup>31</sup> Consistent with the law, future Social Security benefits cannot be borrowed against. However, following a medical loss ( $L_j > 0$ ), if assets fall enough in value then Medicaid will pay the medical costs, thereby ensuring that the household never suffers from negative consumption. As noted above, we assume that households do not face any out-of-pocket medical costs unrelated to long-term care. *Hence, the only correlated costs faced by retirees from a negative health care shock in our model is from the loss of assets above the Medicaid qualification threshold due to spending on long-term care.*

Mathematically, the variable  $Tr_j$  in equation (9) is the sum of all government transfers received (DI, Medicaid, and Social Security) and  $T$  is the total tax rate required to finance those transfers. The value of  $T$  is calculated endogenously to ensure a balanced budget, as discussed below.

#### 4.1.4 Household Optimization Problem

Individuals have preferences for consumption and possibly for leaving bequests, which are time-separable, with a constant relative risk aversion (CRRA) felicity. Most of our simulations assume pure life cycle

---

<sup>28</sup>See Table 20 of Social Security Administration (2012*a*).

<sup>29</sup>Hence, a worker who remains in the high productivity state throughout his or her career will have higher benefits, but a smaller replacement rate on previous earnings. Ideally, we would track each person’s average wage throughout his or her lifetime, but this would require an additional state variable that is computationally costly. Instead, consistent with some other papers, we assign an individual who reaches their final working year the expected benefit conditional on the income earned in that year. However, to accommodate deterministic life cycle factors as well as reduce noise, we run 100,000 simulations, track each person’s average wage and calculate the resulting benefit. We then average the benefits within each of eight income groups in the final working year.

<sup>30</sup>In particular, empirically, many people do not receive full benefits based on the statutory formula due to shorter work histories and other factors. This adjustment, however, had little impact on our numerical results. We present results with unadjusted benefits in Reichling and Smetters (2013).

<sup>31</sup>See BEA NIPA Table 1.1.5 for GDP and Social Security and Medicare Boards of Trustees, 2012 for Social Security spending.

households with no bequest motives, and so any bequests are accidental and come from households that die while holding bonds. But we will also allow for bequest motives as part of our robustness checks. To avoid problems with tractability and uniqueness that arise in models with altruism, bequest motives are modeled as “joy of giving,” meaning that households receive utility based on the size of the bequest that they leave independent of the utility of the recipient:

$$U = \sum_{j=21}^J \beta^j u(c_j) = \sum_{j=21}^J \beta^j \left[ \frac{(c_j - \underline{c})^{1-\sigma}}{1-\sigma} + \xi D_j A_{j+1}^b \right], \quad (10)$$

where  $\beta$  is the subjective discount rate on future utility,  $c_j$  is consumption at age  $j$ ,  $\sigma$  is the risk aversion,  $A_j^b$  is bequeathable wealth held in bonds at age  $j$ ,  $D_j$  is an indicator that equals 1 in the year of death and 0 otherwise,  $\xi$  is a parameter that determines the strength of the bequest motive, and  $J = 120$ , the maximum age. The value  $\underline{c}$  represents minimum subsistence consumption, which we set to zero in our baseline.

An individual’s optimization problem, therefore, is fully described by four state variables: age  $j$ , health  $h$ , idiosyncratic productivity  $\eta$ , and wealth (assets)  $A$ . The household solves the following problem taking the prices  $w$ ,  $r$ ,  $\rho$  as given:

$$V_j(A_j, \eta_j, h_j, j) = \max_{c_j, \alpha_j} \left\{ u(c_j) + \beta s_j(h_j) \int_{h_{j+1}} \int_{\eta_{j+1}} [V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j+1)] Q(\eta_j, d\eta_{j+1}) P_j(h_j, dh_{j+1}) \right\} \quad (11)$$

subject to:

$$A_{j+1} = R(\alpha_j, h_j, h_{j+1})(A_j + X_j - c_j)$$

$$\alpha \leq 1$$

$$\underline{c} \leq c_j \leq A_j + X_j,$$

where:  $A_j = A_j^a + A_j^b$  is beginning-of-period asset consisting of annuities ( $A_j^a$ ) and bonds ( $A_j^b$ );  $R(\alpha_j, h_j, h_{j+1}) = \alpha_j \rho_j(h_j, h_{j+1}) + (1 - \alpha_j)r$  is the portfolio return, where  $\rho_j(h_j, h_{j+1})$  is the annuity return given current health  $h_j$  and future health  $h_{j+1}$  shown in equation (1), which allows for asset rebalancing;  $r$  is the return to the riskless bonds;  $\alpha_j$  is the share of investments made into annuities at age  $j$ ;  $X_j$  is the value of income shown earlier in equation (9).<sup>32</sup> Here,  $A \in \mathbf{R}_+$ ,  $\eta \in \mathbf{D} = \{\eta_1, \eta_2, \dots, \eta_8\}$ ,  $h \in \mathbf{H} = \{h_1, h_2, h_3\}$ ,  $j \in \mathbf{J} = \{21, 22, \dots, 120\}$ , and the functions  $\{V_j, A_j, c_j : \mathbf{S} \rightarrow \mathbf{R}_+\}_{j=21}^{120}$  are measurable with respect to  $\mathbf{F}$ , where  $\mathbf{S} = \mathbf{R}_+ \times \mathbf{D} \times \mathbf{H} \times \mathbf{J}$ ,  $\mathbf{F} = \mathbf{B}(\mathbf{R}_+) \times \mathbf{P}(\mathbf{D}) \times \mathbf{P}(\mathbf{H}) \times \mathbf{P}(\mathbf{J})$ , and  $\mathbf{P}(\cdot)$  denote power sets and  $\mathbf{B}(\mathbf{R}_+)$  is the Borel  $\sigma$ -algebra of  $\mathbf{R}_+$ . Let  $\phi_t(A, \eta, h, j)$  be the population density function of households and  $\Phi_t(A, \eta, h, j)$  be the corresponding cumulative distribution function.

The budget constraints have the following interpretations. Bonds  $A^b$  must be non-negative ( $\alpha \leq 1$ ), thereby recognizing that a competitive market would never allow an individual, who might die before the loan repayment, to borrow at the risk-free rate without also carrying life insurance in the amount of

<sup>32</sup>Individual variables are growth adjusted by  $(1 + \mu)^t$  and aggregate variables are adjusted by  $[(1 + \mu)(1 + \nu)]^{-t}$ , where  $\mu$  is the labor-augmenting growth of productivity and  $\nu$  is the population growth rate.

Table 5: Main Parameter Values

Parameter		Value	Comment
Minimum age		21	
Retirement age		65	Full retirement age
Maximum possible age	$J$	120	
Population growth rate	$\nu$	0.010	Population growth in 1971-2011
Household survival rates	$s_j(h_j)$		Robinson (1996), Social Security Administration (2005; 2013a)
Health transition probabilities	$P_j(m, n)$		Robinson (1996), Social Security Administration (2005; 2012b; 2013a; 2013b), U.S. Census Bureau (2011a)
Long-term care costs	$L$	1.175	Long-term care costs are \$82,670 per year, or 1.2 times average household earnings in the model
Discount factor	$\beta$	0.82-0.98	Discussed in the text
Coefficient of relative risk aversion	$\sigma$	3.000	
Minimum consumption requirement	$\underline{c}$	0.000	Discussed in the text
Share parameter of capital stock	$\omega$	0.320	Nishiyama and Smetters (2005)
Depreciation rate of capital stock	$\delta$	0.046	Nishiyama and Smetters (2005)
Total factor productivity	$\theta$	0.935	Average wage = 1
Productivity growth rate	$\mu$	0.018	Real GDP per capita growth in 1971-2011
Wage process	$\varepsilon_j, \eta_j$		Nishiyama and Smetters (2005)

One unit of earnings in the model economy is equal to \$70,340.

the loan. However, annuities  $A^a$  can potentially be negative ( $\alpha < 0$ ), which is equivalent to borrowing at the risk-free rate and purchasing life insurance to ensure repayment, although much of our simulation analysis below will focus on the non-negative case ( $0 \leq \alpha \leq 1$ ). Moreover, an individual's consumption  $c$  must always remain above the subsistence level  $\underline{c}$ . Without health shocks, this last constraint would never bind under the standard Inada utility conditions. However, with medical expense shocks, we must explicitly enforce the constraint by calculating the required Medicaid payment accordingly.

## 4.2 Production

The production side of our economy is less central in our focus. But as we show below, the level of wealth held by households materially impacts the fraction annuitized, and so we want to ensure that our model produces a plausible capital–output ratio. Moreover, a production side of the economy allows us to recalibrate to the same observable economy when performing sensitivity analysis.

In each period, the representative firm chooses capital  $K$  and labor  $L$  to maximize its period profit,



taking factor prices,  $r_t$  and  $w_t$ , as given,

$$\max_{K_t, L_t} F(K_t, L_t) - (r_t + \delta)K_t - w_t L_t,$$

where  $\delta$  is the depreciation rate. Aggregate output  $F(\cdot)$  is produced by the constant-returns-to-scale production function,

$$F(K_t, L_t) = \theta K_t^\omega L_t^{1-\omega},$$

where  $\theta$  is the total factor productivity and  $\omega$  measures the capital share of output. Capital at time  $t$ ,  $K_t$ , is the sum of bond and annuity holdings by households,

$$K_t = \sum_{j=21}^J \int_{D \times H} (A_t^a(\eta, h, j) + A_t^b(\eta, h, j)) d\Phi_t(A, \eta, h, j).$$

The economy is then described by the measure  $\phi(A, \eta, h, j)$  of individuals by state, and by the values of market wage  $w$ , interest rate  $r$ , capital stock  $K$ , and labor supply  $L$ . Macroeconomic variables are also calibrated consistent with Nishiyama and Smetters (2005). For quick reference, Table 5 gives a summary of the main parameters in our model. The capital share of output is  $\omega = 0.32$ , the depreciation rate of physical capital is  $\delta = 0.046$ , and the capital-to-output ratio is 2.8.<sup>33</sup> The rate of population growth  $\nu$  is assumed to be a constant 1 percent, roughly the value in the United States. The rate of labor-augmenting productivity growth  $\mu$  is set to 1.8 percent, the real GDP per capita growth rate from 1971 to 2011. The capital-output ratio is set at 2.8 by varying the subjective discount rate  $\beta$ , producing a marginal product of capital (interest rate) of  $r = 6.8$  percent.

### 4.3 Payroll taxes

We model Social Security income as a pay-as-you-go transfer from workers to retirees in each period. The Social Security tax rate is determined endogenously under the balanced-budget constraint from the distribution of households in the economy. DI and Medicaid transfers are also financed through a labor income tax. The total tax rate on labor  $T$  is calculated to ensure a balanced budget of all three programs.

### 4.4 General Equilibrium

A general equilibrium is fairly standard, and so a formal definition will be skipped. In particular:

- (i) Household Optimization: Households optimize program (11), taking as given the set of factor prices and policy parameters;
- (ii) Asset Market Clearing: The factor prices are derived from the production technology, with the aggregate levels of saving and labor properly integrated across the measure of households (by Walras' Law, the goods market is redundant and also clears);
- (iii) Policy Balance: The policy parameters are consistent with balanced budget constraints (i.e., tax revenue equals spending); and

<sup>33</sup>This value is at the higher end of the range estimated by Antunes, Cavalcanti and Villamil (2015) who use a baseline capital-output ratio of 2.55, based on the Penn World Tables 6.2. Below, we show that a smaller capital-output ratio reduces annuitization.

Table 6: Age Structure of the Population: Data vs. Model

Age	Data	Model
21-29	17%	20%
30-39	18%	20%
40-49	20%	18%
50-59	19%	16%
65+	18%	19%

Source: U.S. Census Bureau (2011*b*) and authors' calculations. Percentages are based on the population 21 years and older.

(iv) **Bequest Clearing:** Bequests given equal bequests received.

The household's partial equilibrium program (11), therefore, is solved many times, inside of a Gauss-Seidel like iteration, until general equilibrium is reached, defined as having small Euler equation errors away from any boundaries (Judd, 1998). See Appendices E and F for more details.

## 4.5 The Implied Population, Income, and Wealth Distributions

In our baseline model, annuities must be non-negative (no shorting), an assumption that we will relax later. All bequests are accidental ( $\xi = 0$ ) and distributed evenly to surviving households, an assumption that we will relax in the sensitivity analysis. We now examine how the model calibrates to some observable aggregate distributions.

**Population Distributions.** The age structure of our model is fairly similar to 2010 Census data, as shown in Table 6.

**Income Distribution.** Our baseline model calibrates fairly well to the observable data on income inequality. The income Gini coefficient (inclusive of wage income, Social Security, DI, and other benefits) is 0.48 in our model, compared with 0.47 in the data (U.S. Census Bureau, 2012*a*). Moreover, about 9.0% of workers in our model are above the payroll tax ceiling, compared with 8% for males in the data (Social Security Administration, 2012*a*).

**Wealth Distribution.** The amount of wealth inequality (inclusive of bonds and annuities) in our model is below the empirical evidence. The model's wealth Gini coefficient is 0.59, in contrast to the empirical estimate reported by Nishiyama (2002) of 0.75. Table 7 shows the share of wealth held by different wealth percentile groups for both the model and the corresponding empirical evidence from the Census. Almost all of the difference between the model and the data is due to the model's inability to capture the high concentration of wealth held by the top 1%, a gap equal to 20.9% of wealth (9.6% vs. 30.5%), which, in turn, persists throughout the "Top 20%" of the wealth distribution. We narrow the gap somewhat in our sensitivity analysis when we turn on intentional and unequal bequests.

**The Rich.** Life cycle models like ours are notorious for under-predicting the amount of wealth held by the top 1%, likely because they ignore the "entrepreneurial spirit" of the top wealth holders

Table 7: Wealth Distribution: Data vs. Model

	Data <sup>(1)</sup>	Model
Gini Coefficients		
Income <sup>(2)</sup>	0.47	0.48
Wealth	0.75	0.59
Share of Wealth (%)		
Top 1%	30.5	9.6
Top 5%	53.9	26.7
Top 10%	64.9	39.5
Top 20%	77.2	58.4
Top 40%	90.4	85.8
Top 60%	96.9	96.2

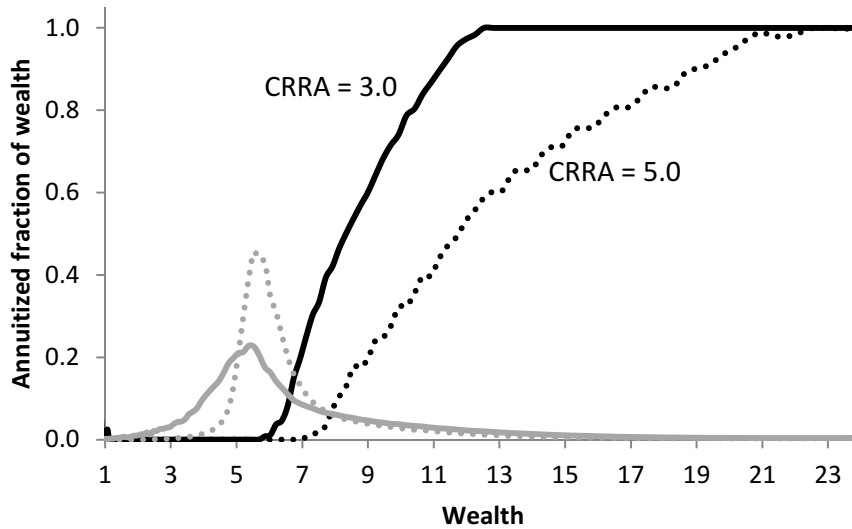
Sources: (1) Nishiyama (2002), (2) U.S. Census Bureau (2012a).

(Cagetti and De Nardi, 2006). Since we are calibrating to an empirically plausible capital-output ratio, less affluent households tend to have more wealth in our baseline economy relative to empirical estimates. Richer households annuitize more of their wealth in our economy, as they are more insulated to health care expense shocks. Our predicted level of annuitization along the *extensive* margin, therefore, is, if anything, biased *upward*, which our sensitivity analysis in Section 5.2.2 demonstrates.

Failing to capture the empirical amounts of wealth held by the top 1%, however, might, in theory, lead us to underpredict the *intensive* margin of annuitization. Still, we believe that any reasonable model of “entrepreneurial spirit,” would only strengthen our key conclusion regarding low optimal annuitization. Because the rich in our model have no entrepreneurial motives, they are attracted mainly to the larger returns offered by the annuity. While entrepreneurs tend to be wealthy, they also have more limited ability to annuitize their wealth. For example, it would be quite challenging to design an annuity wrapper around an individual founder’s privately held equity. To the extent that wealth directly enters an entrepreneur’s utility function in addition to consumption, annuitized wealth also provides less opportunity to create social capital, receive recognition associated with prestige assets, and ability to invest in non-publicly traded firms.

**The Poor.** The total *poverty rate* among *all workers* between the ages of 18 to 64 in our model is 9.0%, which is slightly above the Census value of 7.2% (U.S. Census Bureau, 2012b). Our poverty rate for all *individuals*, including disabled workers, between the ages of 18 to 64 is 10.8%, which is slightly below the Census value of 13.7 percent (U.S. Census Bureau, 2012b). However—and more important for our purposes—the *poverty rate among all disabled people* between the ages of 18 to 64 is 28.6% in our model, very close to the the empirical counterpart of 28.8% estimated by U.S. Census Bureau (2012b) but lower than the value of 50% estimated by Congressional Budget Office (2012a). This number is important because disability during working years produces a correlated shock—comprising lost wages and a reduction in the annuity value—and is one of the drivers for incomplete annuitization. It appears that we are not overestimating its impact.

Figure 7: Annuitized Fraction of Wealth at Age 65 for a Healthy Person



*Explanation:* The black lines show the optimal fraction of wealth held in an annuity by a healthy person ( $h = 1$ ) at age 65 for different coefficients of risk aversion, plotted as a function of wealth. One unit of wealth is equal one year of long-term care costs, or \$82,670. The gray lines with matching plot patterns represent the distribution of households with indicated constant relative risk aversion (CRRA).

*Assumptions:* Annuities are required to be non-negative. Social Security exists. Asset management fees and bequest motives are absent. The capital-output ratio is set to 2.8 by varying the subjective discount rate.

## 5 Simulation Results

We now present simulation evidence using the multi-period model. At key points in the analysis, we compare our results with a similarly calibrated Yaari model where health shocks are turned off and the mortality rate by age is set equal to the average mortality rate in the baseline economy.

### 5.1 Baseline Model

In our baseline model, recall that annuities must be non-negative. We first examine the level of annuitization at age 65 followed by the level of annuitization across the life cycle.

#### 5.1.1 Annuitization at Age 65

The darker lines in Figure 7 show the optimal amount of assets annuitized at the retirement age 65 by a healthy person ( $h = 1$ ) at different levels of wealth achieved by that age. Wealth is reported as a fraction of the average cost of long-term care, around \$83,000 (Section 4.1.3). Notice that the level of annuitization varies significantly by the level of risk aversion  $\sigma$  and wealth. For households with  $\sigma = 3$  in our baseline economy, annuities are not purchased at levels of wealth below 6 times the value of long-term care costs, a total close to a half a million dollars.<sup>34</sup> For households with a higher level

<sup>34</sup>Figure 7 shows a small blip of positive annuitization for households at very low levels of wealth. These households have asset levels close to Medicaid threshold, and so they face only upside potential by annuitizing.

of risk aversion of  $\sigma = 5$ , annuities are not purchased at levels of wealth below 7.5 times the value of long-term care costs. Notice, therefore, that a larger risk aversion  $\sigma$  *decreases* the demand for annuities as aversion grows to the correlation between the remaining value of the annuity and health care costs. In contrast, when we ran the Yaari version of our model (without health shocks), we verified that all age-65 households fully annuitized their assets at all levels of wealth for both  $\sigma = 3$  and  $\sigma = 5$ .<sup>35</sup>

For retirees, annuitization becomes more desirable at larger values of wealth. After a negative survival shock, a wealthy retiree has enough assets to pay for any potentially correlated long-term care cost from the annuity stream itself. Unlike workers, a retiree does not have to worry about any reduction in earnings from becoming disabled or very sick.

However, most new retirees do not have that much wealth. The gray lines with matching patterns in Figure 7 show the percentage of age-65 households with the indicated level of wealth. Both of these distributions fall to left of the point of positive levels of annuitization and far to the left of the points of full annuitization. With  $\sigma = 3.0$ , 44% of healthy retirees at age 65 do not hold annuities and only 7% fully annuitize. Among all retirees, 69% do not hold annuities and only 11% fully annuitize. Recall that these results are not driven by ad-hoc liquidity constraints; agents can fully rebalance their asset portfolio.

While Figure 7 focuses on healthy people ( $h = 1$ ) at age 65, annuitization is also increasing in the health indicator  $h$ . In particular, while only 30% of wealth held by healthy retirees ( $h_1$ ) is annuitized, 61% of all wealth held by retirees with impaired health ( $h_2$ ) is annuitized, increasing to 100% for wealth held by very sick people (state  $h_3$ ). The reason why annuitization increases as health deteriorates is that the upside risk (becoming healthier) and the associated gains in annuity values is increasing, while the downside risk (getting sicker) and the associated losses in annuity values decrease.<sup>36</sup>

### 5.1.2 Comparison with Empirical Data

Table 8 reports the extensive margin of fixed annuity holdings for households age 55 and older across wave 2 (1993) through wave 11 (2012) in the Health and Retirement Study (HRS) panel.<sup>37</sup> Except for wave 1, the HRS contains the standard self-reported health measures (ADLs and IADLs) that we can map into our health states ( $h_1, h_2, h_3$ ), as explained in more detail in Appendix F. We report levels of annuitization sorted by wealth quintile for two objective and broader measures of health: the non-institutionalized and institutionalized (nursing home) populations. For robustness, we present the data using two combinations of sorting. In the first set of columns, we sort households into wealth quintiles and then calculate the fraction of households with any annuities. In the second set of columns, we first sort households by their health measure tuple ( $h_1, h_2, h_3$ ), then group them by wealth quintile within each health category. We then report the weighted average across the wealth quintiles. Since health and wealth are not perfectly correlated, the two measures produce slightly different but broadly consistent results.

Three stylized facts emerge from Table 8. First, empirical annuitization levels are lower than those produced by our baseline model. We return to this issue below where we include more market imperfections that reduce annuitization even more. Second, annuitization rates increase by wealth quintile. Third, annuitization rates are generally higher for the sicker, nursing home population. To be sure, our model is meant as normative rather than positive, especially since we don't believe that the world is

<sup>35</sup>Moreover, when we ran our model with health transitions operative but long-term care costs set at zero, all wealth was again fully annuitized, suggesting that the *impatience channel* investigated in the Appendix is not driving these results either.

<sup>36</sup>Figure 6 shows how annuity returns vary by age and health state transition.

<sup>37</sup>We excluded wave 1 (1992), because that wave does not include data on ADLs and IADLs, the health measures we use to map the data into our model health states.

Table 8: Annuitization in the Data: The Health and Retirement Survey

Sorted by:	Wealth		Health State, then Wealth	
	Home	Nursing Home	Home	Nursing Home
Population:				
Wealth Quintile				
First	1.0%	1.3%	1.1%	1.3%
Second	2.4%	2.2%	2.4%	2.4%
Third	3.9%	4.4%	3.9%	5.2%
Fourth	5.2%	7.6%	4.8%	8.0%
Fifth	4.8%	15.1%	5.0%	13.7%
Total	3.5%	6.1%	3.5%	6.1%

Source: HRS data for wave 2 (1993) through wave 11 (2012); Ages 55+.

Explanation: Shows the fraction of households that have annuities in the HRS data for 1993-2012; Ages 55+. The average sample size per HRS wave for people living at home is about 16,500; the average sample size for people living in nursing homes is 350. Thus, there are roughly 169,000 observations across the 10 waves.

nearly as frictionless as in our baseline economy. But the positive relationships between annuitization and wealth and between annuitization and poor health are broadly consistent with our model’s predictions.

### 5.1.3 Annuitization across the Life Cycle

Across all ages, 38% of wealth is annuitized at  $\sigma = 3.0$ .<sup>38</sup> However, only 26% of all households in the economy hold an annuity; the other 74% hold none. Moreover, only 9% of households fully annuitize. Of course, there is considerable heterogeneity by health status in our baseline calibration: only 21% of healthy ( $h_1$ ) households annuitize, whereas 78% of very sick ( $h_3$ ) households should optimally annuitize. Healthy people, however, greatly outnumber very sick households. We consider additional adjustments in Section 5.2 that reduce annuitization even more.

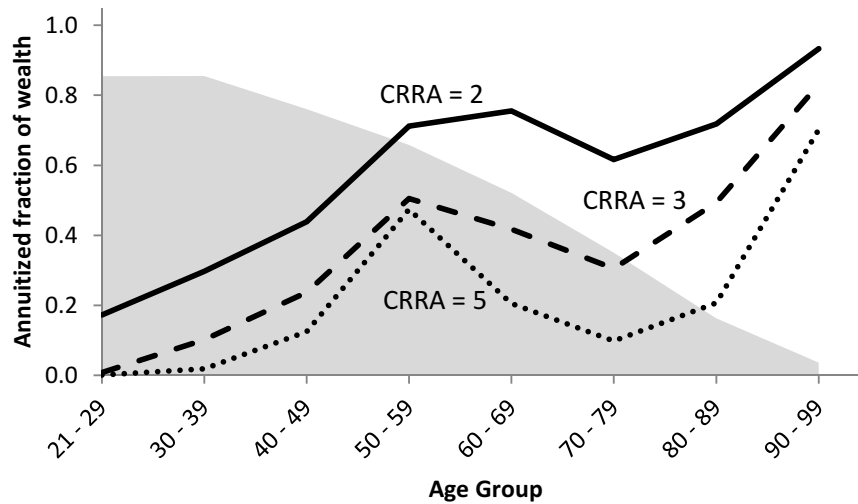
How can 38% of wealth be annuitized when most households do not annuitize at all? The answer lies in the skewness of the wealth distribution. Wealthier households hold a larger fraction of aggregate wealth, and they can more “afford to” annuitize because they can pay for more of the costs associated with negative mortality shocks out of the annuity income stream.

Annuitization across the lifecycle is, however, not monotonic in age. Figure 8 shows the fraction of all wealth that is annuitized (the intensive margin) across the life cycle, while Figure 9 shows the fraction of households that annuitize any assets (the extensive margin). The values presented in both figures are weighted across all health states. The *population density* is shown as a gray shadow and is now independent of the level of risk aversion.

Consider first younger cohorts, which tend to be healthy ( $h_1$ ). The risk of mortality for a healthy younger person is low, and so is the corresponding mortality credit that could be earned from annuitiz-

<sup>38</sup>Table 6 shows that our model slightly over-predicts the population share of younger people and slightly under-predicts the population share of middle-age households. To see if this difference materially impacts our aggregate annuitization levels, we also computed the share of wealth annuitized if we simply exogenously shifted our population weights in order to exactly match their empirical counterpart. Total wealth annuitized increased by about 1 percentage point.

Figure 8: Intensive Margin: Annuitized Fraction of All Wealth, by Age



*Explanation:* Amount of total wealth that is annuitized by age and coefficient of constant relative risk aversion, weighted across all health states. The gray shadow represents the population density.

*Assumptions:* Annuities are required to be non-negative. Social Security exists. Long-term care costs are equal to 1.2 times average annual earnings. Asset management fees and bequest motives are absent. The capital-output ratio is set to 2.8 by varying the subjective discount rate.

ing. The chance of moving from good health to worse health may not be large in absolute terms, but it is *relatively* larger than the value of the small mortality credit. Younger people also have less wealth; as shown above, annuitization rates increase in wealth. Most younger households choose, therefore, not to annuitize. As the age increases, the mortality credit and amount of wealth both increase, leading to more annuitization. However, annuitization begins to fall around age 60 when  $\sigma = 3.0$  and even more sharply at a higher level of risk aversion equal to  $\sigma = 5.0$ . The reason is that as retirement approaches, so does the possibility of uninsured long-term expenses. (Recall that we conservatively assume that workers do not face any uninsured health expenses.) Around age 80, annuitization again begins to climb because the mortality credit quickly grows as the probability of death increases nonlinearly.

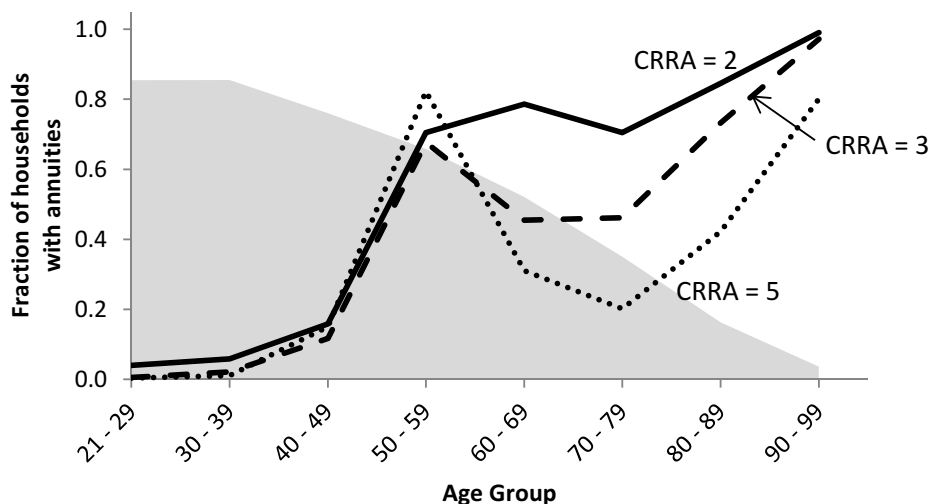
## 5.2 Sensitivity Analysis

We now consider several changes to the baseline model that allow us to investigate the importance of various assumptions that we made. In each case, full annuitization would still exist in the Yaari model.

### 5.2.1 Long-Term Care Expenses

Table 9 investigates the importance of changing uninsured annual nursing home care costs away from our current value of 1.2 times average annual earnings. Smaller values are consistent with the fact that some households (about 7%) secured some form of long-term care protection in the past. Larger values are consistent with the fact that our analysis thus far has ignored other forms of non-insured health care and non-nursing long-term care costs that are likely correlated with the health state; we have also

Figure 9: Extensive Margin: Share of Households with Any Positive Amount of Annuities, by Age



*Explanation:* Fraction of households with any annuities by age and coefficient of constant relative risk aversion, weighted across all health states. The gray shadow represents the population density.

*Assumptions:* Annuities are required to be non-negative. Social Security exists. Long-term care costs are equal to 1.2 times average annual earnings. Asset management fees and bequest motives are absent. The capital-output ratio is set to 2.8 by varying the subjective discount rate.

assumed that, at most, only one member of the household would use nursing care.<sup>39</sup> Not surprisingly, retirees annuitize a larger fraction of their wealth as uninsured long-term care costs are reduced, and they annuitize less as long-term care costs are increased. However, because working households are in the majority and face disability risk, only 40% of households in the economy still want a positive level of annuitization even with no long-term care costs at our baseline relative risk aversion of 3.

### 5.2.2 Targeting the Median Wealth-Income Ratio Instead of the Capital-Output Ratio

Table 10 shows that our model tends to overestimate the median household wealth at key saving ages above age 55. Accordingly, we also ran a simulation where we adjusted the subjective discount rate  $\beta$  so that the median wealth for households in the 55–64 range in the model matched the data. The results are shown in Table 11. Notice that annuitization falls, with only 14% of households holding any annuities, down from 26%. Intuitively, as shown above, households with less wealth hold a larger share of wealth in bonds that are uncorrelated with health care shocks.

<sup>39</sup>We also considered alternative specifications of uninsured health care expenses, including, for example, allowing some expenses to fall in health state  $h_2$ . These variations had very little impact on our results, although typically worked to slightly reduce annuitization.



Table 9: Changing Long-Term Care Costs

Long-Term Care Cost	Relative Risk Aversion					
	2	3	5	2	3	5
	Annuitized Percentage of Total Wealth			Percentage of Households with Annuities		
	<i>All Households</i>					
1.80	51%	22%	15%	30%	16%	16%
1.50	57%	29%	19%	33%	21%	20%
1.20	63%	38%	24%	36%	26%	24%
0.60	74%	59%	39%	43%	38%	33%
0.00	77%	65%	51%	45%	40%	37%
	<i>Retirees Only</i>					
1.80	48%	19%	7%	63%	30%	15%
1.50	55%	25%	9%	68%	39%	20%
1.20	66%	37%	15%	74%	52%	27%
0.60	93%	81%	58%	90%	84%	67%
0.00	100%	100%	100%	93%	92%	89%

*Explanation:* Fraction of wealth annuitized and fraction of households with any annuities, for the entire population and for only retirees, at different levels of long-term care expenses. For long-term care costs of 0.00, all wealth is annuitized but some households have zero wealth, thereby holding no annuities.

*Assumptions:* Annuities are required to be non-negative. Social Security exists. Long-term care costs are a varied fraction of average annual earnings, as indicated. Asset management fees and bequest motives are absent. The capital–output ratio is set to 2.8 by varying the subjective discount rate.

Table 10: Median Wealth by Age: Data vs. Model

Age	2007 SCF	Model
Less than 35	\$12.4	\$19.9
35-44	\$92.4	\$44.5
45-54	\$193.7	\$151.8
55-64	\$266.2	\$386.8
65-74	\$250.8	\$407.8
75 or more	\$223.7	\$334.3

*Sources:* Federal Reserve (2012) and authors' calculations (in thousands).

Table 11: Annuitization with Alternative K/Y Ratios, Intensive and Extensive Margins

	Intensive Margin	Extensive Margin	Fully Annuitized
<i>Baseline Calibration: Targeting a Capital-Output Ratio of 2.8</i>			
Non-retirees	38%	20%	7%
Retirees	37%	52%	15%
Total	38%	26%	9%
<i>Sensitivity: Targeting the Median Wealth of 55-64 Year Olds of \$266,000</i>			
Non-retirees	22%	11%	4%
Retirees	21%	28%	10%
Total	22%	14%	5%
<i>Sensitivity: Targeting a Capital-Output Ratio of 1.6</i>			
Non-retirees	14%	7%	3%
Retirees	16%	21%	8%
Total	15%	9%	4%

*Explanation:* Shows the amount of annuitization in the baseline model with a 2.8 capital-output ratio versus the adjusted model where the median wealth at ages 55-64 is reduced to match the empirical household-level data. Also considers a smaller capital-output ratio of 1.6, corresponding to non-housing wealth.

Even this particular robustness check might still be upward biased for annuitization because our measures of wealth includes housing. To the extent that housing wealth cannot be easily borrowed against or reverse mortgaged, a smaller capital-output ratio might be viewed as appropriate. Hence, Table 11 also reports annuitization levels for a capital-output ratio at 1.6 that excludes housing wealth. Annuity rates fall even more.

### 5.2.3 Minimum Required and Guaranteed Consumption

In our baseline calibration, we conservatively set the required level of consumption for subsistence,  $\underline{c}$ , equal to 0.<sup>40</sup> The tradition of Stone-Geary utility, however, recognizes that households enjoy additional utility only once a positive subsistence level of consumption, required for survival, is achieved. Moreover, Ameriks et al. (2011) carefully show empirically that older households have an additional aversion to public care facilities, in particular, Medicaid. Their results suggest that the effective value of  $\underline{c}$  should be higher.

At the same time, our baseline calibration also assumes that while Medicaid covers nursing home expenses to avoid negative wealth, only a small amount (about \$70) of additional consumption is then guaranteed after such a large shock. However, the Medicaid program has another feature called the Personal Needs Allowance (PNA), which is a monthly amount of money that residents who receive Medicaid may retain from their personal income. Any income above the PNA must be paid toward the cost of their care. This allowance is intended for residents “to spend at their discretion on items such as telephone expenses, cigarettes, a meal out with friends, cards to send to family, reading materials, or

<sup>40</sup>To ensure that consumption never actually drops this far in the presence of the utility Inada condition, Medicaid is activated whenever consumption would otherwise drop below \$70, or 0.001 in our model.

hobbies.” (National Long-Term Care Ombudsman Resource Center, 2009) The PNA average across the U.S. states is about \$600 per year (National Long-Term Care Ombudsman Resource Center, 2009).

Table 12 shows how sensitive our results are to changes in the required and guaranteed amount of consumption. (For ease of comparison, the first two rows of values repeat the results from the baseline calibration and when we targeted the median level of wealth between ages 55 and 64.) When the guaranteed minimum consumption is increased to \$600 while conservatively assuming a zero required subsistence level of consumption,  $\underline{c} = 0$ , 22% of households hold at least some annuities and 37% of all wealth is annuitized.

As additional sensitivity analysis, we now consider the empirical estimates by Ameriks et al. (2011). Recall that we have conservatively assumed that at most one person in each household will potentially need nursing care even though this person has access to the household’s joint resources. In contrast, the unit of analysis in Ameriks et al. (2011) is the single retiree rather than a married household. They estimate that single retirees ages 65 and older have a median level of wealth equal to \$88,000 and enjoy a minimum level of consumption equal to \$2,200.

Table 12, therefore, considers a range of parameterizations consistent with their estimates. The fourth row of values shows the level of annuitization for a guaranteed minimum level of consumption equal to \$2,200 and a zero subsistence level,  $\underline{c} = 0$ , while continuing to assume this person has access to the household’s joint resources. Notice that the fraction of households with any annuities increases to 33%, while the fraction of wealth annuitized increases to 53%—a sizeable rise.

But a zero subsistence level of consumption is, of course, not plausible. So, instead, consider an increase in the subsistence level to  $\underline{c} = \$1,600$ . This value is equal to the \$2,200 estimate by Ameriks et al. (2011) less the \$600 PNA. In other words, the first \$1,600 of the \$2,200 is needed for basic survival, with the PNA allowing for additional utility. This interpretation is consistent with the intent of the PNA: to allow for only small amounts of consumption beyond the basic level needed for survival. Now, only 23% of households annuitize any wealth and 37% of all wealth is annuitized. Moreover, as noted above, the analysis by Ameriks et al. (2011) shows that households are especially risk averse to public care facilities, which would argue for an even larger value of  $\underline{c}$ . Such a calibration would produce even lower levels of annuitization.

Finally, consider a wealth value of \$88,000, as estimated by Ameriks et al. (2011) for single retirees ages 65 and older (the third panel). In effect, we are now allowing for both members of our household to potentially need nursing care. With a subsistence level of consumption equal to  $\underline{c} = \$1,600$ , only 13% of households annuitize any wealth and 19% of wealth is annuitized. As before, if we incorporated an additional risk aversion to public care facilities, annuitization rates would fall even more.

Table 12: Effects of Changes in the Required and Guaranteed Levels of Consumption

Minimum Consumption Requirement, $\underline{c}$		Annuitized fraction of wealth			Fraction of households with annuities		
		Total	Retirees only	Non-retirees only	Total	Retirees only	Non-retirees only
<i>Baseline Calibration: Targeting a Capital-Output Ratio of 2.8</i>							
\$0	\$70	38%	37%	38%	26%	52%	20%
<i>Sensitivity: Targeting the Median Wealth of 55-64 Year Olds of \$266,000</i>							
\$0	\$70	22%	21%	22%	14%	28%	11%
\$0	\$600	37%	42%	34%	22%	50%	15%
\$0	\$2,200	53%	70%	46%	33%	80%	23%
\$1,600	\$2,200	35%	41%	33%	22%	49%	16%
<i>Sensitivity: Targeting the Median Wealth of 65+ Year Olds of \$88,000</i>							
\$0	\$2,200	40%	63%	33%	25%	58%	17%
\$1,600	\$2,200	17%	29%	13%	11%	26%	8%

*Explanation:* Shows how varying the amount of required and guaranteed consumption affects the level of annuitization in the baseline model calibrated to a capital-output ratio of 2.8 and alternative calibration targets matching a) the median wealth of families with 55-64 year old heads of household in the 2007 SCF (\$266,000); and b) the median wealth Ameriks et al. (2011) find in their sample of retired singles ages 55 and older (\$88,000).

## 5.2.4 Partial-Equilibrium Calculations and the Rate of Time Preference

Recall that in the Yaari model, the annuitized share of wealth is independent of the level of wealth. In our model, however, the calibrated level of wealth is very important for estimating optimal levels of annuitization since wealthier households annuitize a larger fraction of their resources. In our baseline calibration, achieving a capital-output ratio of 2.8 requires a value of  $\beta$  equal to 0.82 (short sales prohibited), 0.85 (short sales allowed), or 0.90 (the Yaari case). Like with many general-equilibrium lifecycle models where households face idiosyncratic shocks, precautionary motives generate a significant amount of saving, requiring less reliance on intertemporal substitution. Our calibrated values also fall in the range of empirical estimates using field evidence (Hausman, 1979) and experimental evidence (Thaler, 1991). Warner and Pleeter (2001) provide estimates of personal discount rates using a military drawdown program where over 65,000 separatees were offered the choice between an annuity and a lump-sum payment. Despite break-even discount rates exceeding 17 percent, most of the separatees selected the lump sum. In their critical review of the extensive literature on time discount rates, Frederick, Loewenstein and O'Donoghue (2002) write that the Warner and Pleeter (2001) "study is particularly compelling in terms of credibility of reward delivery, magnitude of stakes, and number of subjects." (P. 385). The results of their study are consistent with our range of values for  $\beta$ .<sup>41</sup>

Still, some previous partial-equilibrium simulation studies have used values of  $\beta$  closer to 0.95 and even higher, although they focus on a different mechanism influencing annuitization rates. Table 13 explores this distinction in more detail. For ease of comparison, the first panel ("Baseline Calibration") summarizes the baseline levels of annuitization presented earlier with a 2.8 capital-output ratio. As reported earlier, 37% of all retiree-held wealth is annuitized, 52% hold any annuities, and 15% fully

<sup>41</sup>Frederick, Loewenstein and O'Donoghue (2002) also write that it is challenging to distinguish between intertemporal substitution and behavioral biases, a potentially important issue that we leave for future study.

Table 13: Annuitization in Partial Equilibrium and a Higher Discount Factor

	Intensive Margin	Extensive Margin	Fully Annuitized	K/Y Ratio
<i>Baseline Calibration (<math>\beta = 0.82</math>)</i>				
Non-retirees	38%	21%	7%	
Retirees	37%	52%	15%	
Total	38%	26%	9%	2.8
<i>Partial Equilibrium with <math>\beta = 0.98</math></i>				
Non-retirees	78%	57%	34%	
Retirees	90%	97%	62%	
Total	82%	65%	39%	6.1
<i>Partial Equilibrium with <math>\beta = 0.98</math> and Baseline Wealth Distribution</i>				
Non-retirees	51%	29%	13%	
Retirees	41%	52%	17%	
Total	48%	34%	14%	2.8

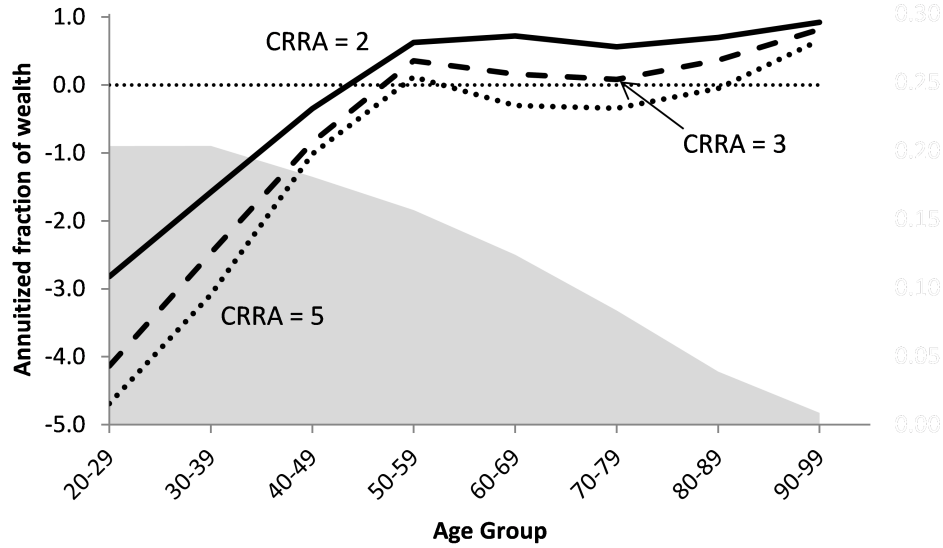
*Explanation:* Shows the amount of annuitization in the baseline model with  $\beta = 0.82$  versus a partial-equilibrium calibration with  $\beta = 0.98$  and a “hybrid” partial-equilibrium simulation where we exogenously fix the wealth distribution across households consistent with that in the baseline economy, but using the decision rules of people from the partial-equilibrium calibration with  $\beta = 0.98$ .

annuitize. The second panel shows the values for a partial-equilibrium calibration where the value of  $\beta$  is increased to 0.98, on the high side of a plausible range. Notice that the capital-output ratio increases dramatically, from 2.8 to 6.1. The corresponding median wealth for households in the 55–64 age range skyrockets to \$650,000, around 240% larger than the corresponding empirical value reported in Table 10. Notice that 90% of retiree wealth is now annuitized, 97% hold some annuities, and 62% fully annuitize. Intuitively, most households in the economy are now quite wealthy, and, as shown earlier, the rich optimally annuitize more of their assets. The third panel shows a “hybrid” partial-equilibrium simulation where we exogenously fix the wealth distribution across households consistent with that in the baseline economy, producing a 2.8 capital-output ratio, but then examine what decisions people would make with  $\beta = 0.98$ .<sup>42</sup> Notice that only 41% of retiree wealth is now annuitized (compared to 37% in the baseline), 52% hold any annuities (identical to the baseline), and 17% fully annuitize (compared to 15% in the baseline). In sum, our assumed rates of time preference play only a small role in impacting annuitization, even under a fairly stylized construction that is most favorable to its role. In fact, when we completely turn off correlated uninsured health care costs, we obtain 100% annuitization for the entire range of values of  $\beta$  in our baseline model, with or without a short sales constraint.<sup>43</sup>

<sup>42</sup>In particular, we solve the general-equilibrium with  $\beta = 0.98$  to get a consistent solution but then reallocate the population measures onto the wealth grid points consistent with the baseline model.

<sup>43</sup>In models *with* liquidity constraints, a small value of  $\beta$  can produce incomplete annuitization if people want a decreasing real path of consumption in retirement (Davidoff, Brown and Diamond, 2005). This effect, however, is absent in our model where households can rebalance their level of annuitization. Instead, as Reichling and Smetters (2013) shows, the value of  $\beta$  can impact the optimal design of the annuity contract, and a level annuity might be suboptimal at low values of  $\beta$ . However, as noted in the text, in the simulations reported in this paper, all assets are annuitized if correlated uninsured health care costs

Figure 10: Short Sales: Annuitized Fraction of All Wealth, by Age (Intensive Margin)



*Explanation:* Amount of total wealth that is annuitized by age, weighted across all health states. The gray shadow represents the population density.

*Assumptions:* Negative annuity holdings are allowed. Social Security exists. Long-term care costs are equal to 1.2 times average annual earnings. Asset management fees and bequest motives are absent. The capital-output ratio is set to 2.8 by varying the subjective discount rate.

### 5.2.5 Allowing for Short Selling

Thus far, our simulation analysis has restricted annuity allocations to be non-negative. That assumption is not strictly necessary. Although a private market would not lend bonds at the risk-free rate because agents can die, annuities could go negative provided that bonds are positive.<sup>44</sup> As discussed earlier, a negative annuity is equivalent to purchasing life insurance.

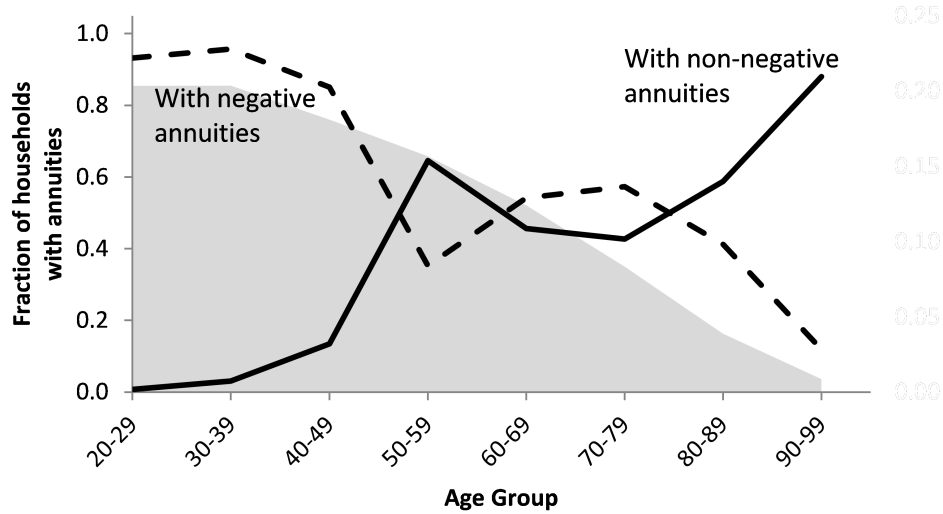
Figure 10 shows the amount of total wealth annuitized by age group, again with no intentional bequest motives ( $\xi = 0$ ). Notice that younger households, which tend to be healthy, hold a negative (short) position in annuities. On one hand, a negative position is costly to younger households because they must now *pay* for the mortality credit. But the mortality credit is also relative inexpensive when young. On the other hand, the negative annuity position provides a valuable hedge against future negative health shocks that could otherwise reduce their income before retirement and/or produce expenses after retirement. Specifically, after a future realization of negative health information, this short position can be reversed by going long in an annuity that is now less expensive than before the negative health shock. The difference in the value of these short-long trades produces a net profit to the household that can be used to pay for any correlated income loss and/or uninsured expenses.

Figure 11 shows in more detail the fraction of households with positive or negative annuities. Notice that, within the same age group, some households might hold a positive level of annuities and others

are eliminated.

<sup>44</sup>The requirement of positive bonds reflects our requirement that people do not have negative total wealth. That restriction would not be binding in the presence of the Inada condition and risky wage income if there were not safe sources of income, including Social Security and disability. We are effectively preventing people from borrowing against those sources.

Figure 11: Short Sales: Share of Households with Any (Positive or Negative) Amount of Annuities, by Age (Extensive Margin)



*Explanation:* Fraction of households with any (positive or negative) annuities by age, weighted across all health states. The gray shadow represents the population density.

*Assumptions:* Negative annuity holdings are allowed. Social Security exists. Long-term care costs are equal to 1.2 times average annual earnings. Asset management fees and bequest motives are absent. The capital–output ratio is set to 2.8 by varying the subjective discount rate. The value of  $\sigma = 3.0$ .

might have negative holdings. This fairly striking pattern is due to the heterogeneity in our model. Households of the same age can vary in health status, income realization, and the amount of inherited assets *relative* to their permanent income. Notice that even some older households want a negative annuity position, including almost half of households at age 65.<sup>45</sup>

Of course, a short position appears to be contrary to conventional wisdom and practice. Only about 17% of individuals between ages 18 and 24 hold individual life insurance policies, which increases to about 26% for people between the ages of 25 and 34 (LIMRA, 2011).<sup>46</sup> Moreover, the apparent primary motivation for buying life insurance is to protect dependents rather than to hedge future health risks (LIMRA, 2012b). In contrast, in our model, it is optimal for most younger households to short annuities (purchase life insurance), even if they have no dependents or bequest motives. Hence, in practice, households might be making ill-informed choices or narrowly framing their decisions, consistent with Brown et al. (2008) and Beshears et al. (2012). Or, the conventional guidance given to households could simply be suboptimal. We leave that reconciliation to future research. Still, we modestly suggest that the “true annuity puzzle” might actually be why we do not see more *negative* annuitization (life insurance) by younger households.

<sup>45</sup>The non-monotonic relationship between shorting and age reflects interactions between the rate of time preference and the evolution of health status over the lifecycle. Younger workers are more concerned about lost wages and early retirees are more concerned with nursing home expenses.

<sup>46</sup>If group life insurance policies are included, these numbers increase to 36% for ages 18–24 and 54% for ages 25–34.

## 5.2.6 State Dependent Utility

So far, our analysis has ignored the relationship between marginal utility and health status. However, Scholz and Seshadri (2012) estimate that consumption and health are complements. Similarly, Finkelstein, Luttmer and Notowidigdo (2013) find that the marginal utility of consumption declines as health deteriorates. They estimate that a one-standard-deviation increase in an individual's number of chronic diseases is associated with a 10%-25% decline in marginal utility among people age 50 and older. Peijnenburg, Nijman and Werker (2013) finds that state dependent utility has little impact on annuitization rates in the liquidity-constraints model. We show that these results extend to our correlated-risk model, where the value of the annuity falls after a negative health shock.

In particular, to analyze the impact of state dependent utility, we now modify the felicity function  $u(c_j)$  shown in equation (10) to equal

$$u(c, h) = \frac{(c - \underline{c})^{1-\sigma}}{1-\sigma} h^\lambda,$$

where we turn off bequest motives (i.e., set  $\xi = 0$ ).<sup>47</sup> The term  $h$  is health, where, recall, larger values represent worsening health. The marginal utility with respect to consumption  $c$  is then equal to  $u_c(c, h) = (c - \underline{c})^{-\sigma} h^\lambda$ . This marginal utility which changes in  $h$  according to  $u_{c,h}(c, h) = \lambda (c - \underline{c})^{-\sigma} h^{\lambda-1}$ . For  $c$  and  $h$  to be complements (i.e.  $u_{c,h}(c, h) < 0$ , since health is decreasing in  $h$ ) the value of  $\lambda < 0$ .

Table 14 shows how the annuitization decisions of retirees along the intensive and the extensive margins change with different parameter values of  $\lambda$ , where  $\lambda = 0$  corresponds to our baseline economy. The parameter  $\lambda$  determines the strength of the complementarity between consumption and health, and thus, by how much consumption decreases as households move from being healthy to impaired (the  $h_1 \rightarrow h_2$  transition) or from being impaired to disabled (the  $h_2 \rightarrow h_3$  transition). The values of  $\lambda$  are chosen based on the estimates provided by Finkelstein, Luttmer and Notowidigdo (2013) mentioned above.<sup>48</sup> At the lower bound of estimates, the marginal utility of consumption,  $u_c$ , decreases by 16% when retirees transition from  $h_1$  to  $h_2$  and by 10% when retirees transition from  $h_2$  to  $h_3$ . The associated increase in annuitization, compared to our baseline, is 2 percentage points along the intensive margin and 3 percentage points along the extensive margin. Changes in annuitization are still modest for estimates of  $\lambda$  at the upper end of effects that Finkelstein, Luttmer and Notowidigdo (2013) find. The decreases in the marginal utility are more than twice as large when  $\lambda = -0.77$  compared to the case with  $\lambda = -0.26$  and retirees annuitization increases by 5 percentage points along the intensive margin and 8 percentage points along the extensive margin, as compared to the baseline.

## 5.3 Additional Factors That Reduce Annuity Demand

Thus far, our analysis has considered the role of uninsured expenses that reduce the level of annuitization in our model with stochastic mortality probabilities but leave the level of annuitization at 100% with deterministic mortality probabilities. Relative to our model, annuity prices in the real world are

<sup>47</sup>We purposely keep the modification as simple as possible for sake of comparison. If people could additionally make an investment in health, a more general formulation would be more appropriate.

<sup>48</sup>We map Finkelstein et al.'s estimate (which is based on the number of chronic conditions) into our health states (which are defined based on IADL and ADL impairments) by assigning people sampled in the Health and Retirement Study to health states  $h_1$  through  $h_3$  based on their reported IADL and ADL impairments. We then use the average number of chronic conditions per health state as the basis to estimate how a change in the health state would affect the number of chronic conditions people have (on average). That change in the average number of health states together with Finkelstein et al.'s estimates are then translated into a range of estimates for the parameter  $\lambda$ . For a detailed discussion see Appendix F.



Table 14: Effects of Changes in  $\lambda$ , Retirees Only

$\lambda$	$\Delta$ in $u_c$		Annuitization of Retirees	
	$h_1 \rightarrow h_2$	$h_2 \rightarrow h_3$	Intensive Margin	Extensive Margin
0.00	0%	0%	37%	52%
-0.26	-16%	-10%	39%	55%
-0.49	-29%	-18%	40%	57%
-0.77	-41%	-27%	42%	60%

*Explanation:* Shows how the annuitization decisions of retirees change with changes in the complementarity between consumption and health,  $\lambda$ .  $\lambda = 0$  corresponds to the baseline economy. A more negative  $\lambda$  indicates greater complementarity between consumption and good health. The values of  $\lambda$  are chosen such that the changes in the marginal utility of consumption  $u_c$  are consistent with estimates by Finkelstein, Luttmer and Notowidigdo (2013).

more expensive, likely due to the presence of sales and risk loads, management fees and asymmetric information.<sup>49</sup> According to IncomeSolutions.com, a large online broker of fixed annuities, it costs a healthy 65-year old \$178,500 for an immediate annuity that pays \$1,000 per month; in our model, the cost for a healthy person at age 65 is only \$98,000.<sup>50</sup>

Appendix G considers additional factors that can reduce annuitization in order to see how far a rational model can be pushed to produce a low level of annuitization. To be conservative, we continue to target a 2.8 capital-output ratio rather than lower levels of aggregate wealth. Overall, with realistic levels of management fees and bequests, less than 10% of households hold any annuities, and even a smaller fraction of wealth is annuitized. These results are robust to a variety of ways that bequests might be transmitted across generations. Appendix G argues that additional model modifications – differential transaction costs, more worker risk, and asymmetric information – would reduce annuitization even more.

Another modification is if households believe that the annuity provider could potentially default, a second important factor found in the survey by Beshears et al. (2012) for why households might not annuitize.<sup>51</sup> Even very small perceived default risks can have a large impact on the demand for annuities because the probability of default falls outside of the felicity function whereas the value of the mortality credit on consumption from annuitization falls inside. As a result, the mortality credit must be much larger than the probability of default for a risk-averse household to heavily annuitize. With zero management fees and no bequest motive, the share of households that hold no annuities increases slightly from 74% to 77% if the probability of default is just 1 in 1,000 years. The share of wealth annuitized, however, falls from 38% to 23%, and only 2% of households fully annuitize, down from

<sup>49</sup>Recall that in the Yaari model, asymmetric information should not reduce annuitization. With correlated cost shocks, as in our model, asymmetric information can reduce annuitization. However, to be conservative, we have assumed throughout that insurers and households have access to the same information.

<sup>50</sup>Last Checked: October 23, 2014. In more recent times, annuity providers might be using lower discount rates as well. For example, at a discount rate of 3.0%, the cost of our annuity increases to \$151,537, although still less than the \$178,500 commercial price.

<sup>51</sup>As Beshears et al. (2012) note, state guarantee funds exist against such a default, although the limits are capped and bailouts are policy rather than a contract. Hence, we can think of the small perceived default probability that we simulate as a combination of the insurer defaulting and the state not paying. To be clear, we are not arguing that such beliefs are rational. Rather, given the belief, it can have a large impact on the demand for annuitization.

9%. We have also performed “kitchen sink” simulations (not reported) where we turn on several of these frictions at the same time, thereby achieving even lower levels of annuitization on the intensive and extensive margins.

## 6 Conclusions

This paper shows that the original Yaari prediction of 100% annuitization of wealth is very hard to break with various market frictions when survival probabilities are assumed to be deterministic, as is standard. Allowing for the survival probabilities themselves to be stochastic due to changes in health status and including their concomitant costs, however, can break the full annuitization result and become a gateway mechanism for various market frictions to matter. Our simulation evidence suggests that, even under conservative assumptions, it is indeed not optimal for most households to annuitize any wealth; many younger households should actually *short* annuities. Future work can extend our analysis in several directions as discussed in Appendix G. We believe that most extensions would only decrease the optimal level of annuitization even more. Future work could also examine how the results impact the optimal construction of tax and social insurance policies (Netzer and Scheuer, 2007).

Recovering the standard 100% annuitization result is a fairly tall order when survival probabilities are stochastic and people face correlated longevity costs. In particular, even ignoring transaction costs, bequest motives, default risk and related frictions, several factors must all hold approximately<sup>52</sup> for full annuitization: (i) disability insurance replaces all lost wages; (ii) out-of-pocket medical costs, nursing home costs, and non-nursing long-term care costs are eliminated; (iii) agents are sufficiently patient; and (iv) there is no asymmetric information. All four of these factors are irrelevant in the Yaari model, but each is quite important when survival probabilities are stochastic. Although some amounts of imperfect insurance will likely always be optimal in the presence of private information, tackling each factor in more detail represents an opportunity for future research.

For example, the provocative policy paper by Warshawsky, Spillman and Murtaugh (2002) argues for the creation of a “medical annuity” product that integrates annuity protection with long-term care cost insurance. On one hand, such a product could reduce *adverse selection* since long-lived individuals who raise annuity costs may also produce fewer long-term care costs in present value. (However, partial annuitization would still exist due to worker disability.) On the other hand, it is unclear if consumers would be comfortable with a potential (or just perceived) *moral hazard* problem associated with buying long-term care insurance from a firm that subsequently profits from the consumer enjoying a shorter lifespan.

## References

- A.M. Best Co. 2014. “Best’s Impairment Rate and Rating Transition Study—1977 to 2013.” <http://goo.gl/90k8g7>.
- Ameriks, John, Andrew Caplin, Steven Laufer, and Stijn Van Nieuwerburgh. 2011. “The Joy of Giving or Assisted Living? Using Strategic Surveys to Separate Public Care Aversion from Bequest Motives.” *Journal of Finance*, 66(2): 519–561.

---

<sup>52</sup>We write “approximately” because the mortality credit provides some room for frictions.

- Antunes, Antonio, Tiago Cavalcanti, and Anne Villamil. 2015. "The effects of credit subsidies on development." *Economic Theory*, 58(1): 1–30.
- Auerbach, Alan J., Jagadeesh Gokhale, Laurence J. Kotlikoff, John Sabelhaus, and David N. Weil. 1995. "The Annuitization of Americans' Resources: A Cohort Analysis." National Bureau of Economic Research Working Paper 5089.
- Autor, David H., Nicole Maestas, Kathleen J. Mullen, and Alexander Strand. 2015. "Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants." National Bureau of Economic Research Working Paper 20840.
- Baldwin, William. 2013. "Dodge the Long-Term Care Insurance Mess." *Forbes*, <http://goo.gl/Xv2Dl>.
- Bauer, Daniel, and Jochen Russ. 2012. "A New Methodology for Measuring Actual to Expected Performance." Georgia State University and Institut für Finanz- und Aktuarwissenschaften Working Paper.
- Bernheim, B. Douglas. 1991. "How Strong Are Bequest Motives? Evidence Based on Estimates of the Demand for Life Insurance and Annuities." *Journal of Political Economy*, 99(5): 899–927.
- Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Stephen P. Zeldes. 2012. "What Makes Annuitization More Appealing?" National Bureau of Economic Research Working Paper 18575.
- Bodie, Zvi. 2003. "Thoughts on the Future: Life-Cycle Investing in Theory and Practice." *Financial Analysts Journal*, January/February: 24–29.
- Brown, Jeffrey R., and Amy Finkelstein. 2008. "The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market." *American Economic Review*, 98(3): 1083–1102.
- Brown, Jeffrey R., and Amy Finkelstein. 2011. "Insuring Long-Term Care in the United States." *Journal of Economic Perspectives*, 25(4): 119–42.
- Brown, Jeffrey R., Jeffrey R. Kling, Sendhil Mullainathan, and Marian V. Wrobel. 2008. "Why Don't People Insure Late-Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle." *American Economic Review*, 98(2): 304–09.
- Brugiavini, Agar. 1993. "Uncertainty Resolution and the Timing of Annuity Purchases." *Journal of Public Economics*, 50(1): 31–62.
- Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy*, 114(5): 835–870.
- Centers for Disease Control and Prevention. 2009. "The National Nursing Home Survey: 2004 Overview." *Vital and Health Statistics*, 13(167).
- Centers for Medicare and Medicaid Services. 2013. "Medigap Policy Search." <http://goo.gl/xKL5q>.
- Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. "A Theory of Financing Constraints and Firm Dynamics." *Quarterly Journal of Economics*, 121(1): 229–265.
- CNN Money. 2013. "Annuities—What are the disadvantages?" *Ultimate Guide to Retirement*, <http://goo.gl/j2uG7>.

- Congressional Budget Office. 2012a. “Policy Options for the Social Security Disability Insurance Program.” <http://goo.gl/22wNo>.
- Congressional Budget Office. 2012b. “The Social Security Disability Insurance Program - Infographic.” <http://goo.gl/YEc301>.
- Davidoff, Thomas, Jeffrey R. Brown, and Peter A. Diamond. 2005. “Annuities and Individual Welfare.” *American Economic Review*, 95(5): 1573–1590.
- De Nardi, Mariacristina, Eric French, and John B. Jones. 2010. “Why Do the Elderly Save? The Role of Medical Expenses.” *Journal of Political Economy*, 118(1): 39–75.
- Employee Benefit Research Institute. 2013. “Amount of Savings Needed for Health Expenses for People Eligible for Medicare: More Rare Good News.” *Notes*, 34 (10): 2 – 7.
- Federal Reserve. 2012. “Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances.” *Federal Reserve Bulletin*, 98 (2): 1 – 80.
- Fidelity Benefits Consulting. 2013. “Fidelity Estimates Couples Retiring In 2013 Will Need \$220,000 To Pay Medical Expenses Throughout Retirement.” *Inside Fidelity*. <https://goo.gl/SiydDY>.
- Finkelstein, Amy, and James Poterba. 2004. “Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market.” *Journal of Political Economy*, 112(1): 183–208.
- Finkelstein, Amy, Erzo Luttmer, and Matthew Notowidigdo. 2013. “What Good is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption.” *Journal of the European Economic Association*, 11: 221 – 258.
- Frederick, Shane, George Loewenstein, and Ted O’Donoghue. 2002. “Time Discounting and Time Preference: A Critical Review.” *Journal of Economic Literature*, 40(2): 351–401.
- French, Eric. 2005. “The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour.” *Review of Economic Studies*, 72(2): 395–427.
- Gale, William G., and John Karl Scholz. 1994. “Intergenerational Transfers and the Accumulation of Wealth.” *Journal of Economic Perspectives*, 8(4): 145–160.
- Genworth Financial. 2012. “Genworth 2012 Cost of Care Survey—Executive Summary.” <http://goo.gl/mjalig>.
- Gottlieb, Daniel. 2012. “Prospect Theory, Life Insurance, and Annuities.” The Wharton School Working Paper.
- Hausman, Jerry A. 1979. “Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables.” *Bell Journal of Economics*, 10(1): 33–54.
- Hendricks, Lutz. 2002. “Intended and Accidental Bequests in a Life-Cycle Economy.” Department of Economics, Arizona State University Working Paper.
- Judd, Kenneth L. 1998. *Numerical Methods in Economics*. The MIT Press, Cambridge, Massachusetts.

- Kotlikoff, Laurence J., and Avia Spivak. 1981. "The Family as an Incomplete Annuities Market." *Journal of Political Economy*, 89(2): 372–91.
- Lieber, Ron. 2010. "Ignore Long-Term Care Planning at Your Peril." *New York Times*, <http://goo.gl/2dOP8>.
- Life Insurance Settlement Association. 2013. "Life Settlement History." <http://goo.gl/bb30r>.
- LIMRA. 2011. "Person-Level Trends in U.S. Life Insurance Ownership." *report*, Windsor, Connecticut.
- LIMRA. 2012a. "Annuity Sales Struggle under Weight of Low Interest Rates." <http://goo.gl/fj8gTh>.
- LIMRA. 2012b. "Facts from LIMRA: Life Insurance Awareness Month, September 2012." <http://goo.gl/ezGATx>.
- Lockwood, Lee M. 2013. "Incidental Bequests: Bequest Motives and the Choice to Self-Insure Late-Life Risks." Northwestern University Working Paper.
- Lopes, Paula, and Alexander Michaelides. 2007. "Rare events and annuity market participation." *Finance Research Letters*, 4(2): 82–91.
- Low, Hamish, and Luigi Pistaferri. 2010. "Disability Risk, Disability Insurance and Life Cycle Behavior." National Bureau of Economic Research Working Paper 15962.
- Maurer, Raimond, Olivia S. Mitchell, Ralph Rogalla, and Vasily Kartashov. 2013. "Lifecycle Portfolio Choice with Systematic Longevity Risk and Variable Investment-Linked Deferred Annuities." *Journal of Risk and Insurance*, 80(3): 649–676.
- MetLife. 2012. "Market Survey of Long-Term Care Costs." November. <https://goo.gl/2DetGY>.
- Modigliani, Franco. 1986. "Life Cycle, Individual Thrift, and the Wealth of Nations." *American Economic Review*, 76(3): 297–313.
- National Long-Term Care Ombudsman Resource Center. 2009. "Personal Needs Allowances for Residents Long-Term Care Facilities: A State by State Analysis." <http://goo.gl/g7HN2g>.
- Netzer, Nick, and Florian Scheuer. 2007. "Taxation, Insurance, and Precautionary Labor." *Journal of Public Economics*, 91(7-8): 1519–1531.
- Nishiyama, Shinichi. 2002. "Bequests, Inter Vivos Transfers, and Wealth Distribution." *Review of Economic Dynamics*, 5(4): 892–931.
- Nishiyama, Shinichi, and Kent Smetters. 2005. "Consumption Taxes and Economic Efficiency with Idiosyncratic Wage Shocks." *Journal of Political Economy*, 113(5): 1088–1115.
- Ohlemacher, Stephen. 2013. "Social Security Chief: Program Hurt by Neglect." *Associated Press*, <http://goo.gl/qJn2x>.
- Palumbo, Michael G. 1999. "Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle." *Review of Economic Studies*, 66(2): 395–421.
- Peijnenburg, Kim, Theo Nijman, and Bas Werker. 2013. "Health Cost Risk: A Potential Solution to the Annuity Puzzle." *Working Paper*.

- Prudential. 2010. “Long-Term Care Cost Study.” <http://goo.gl/MmhUNH>.
- Prudential. 2011. “Long-Term Care Insurance: A Piece of the Retirement & Estate Planning Puzzle.” <https://goo.gl/DrVsI1>.
- Reichling, Felix, and Kent Smetters. 2013. “Optimal Annuitization with Stochastic Mortality Probabilities.” *NBER Working Paper No. 19211*.
- Robinson, Jim. 1996. “A Long-Term Care Status Transition Model.” In *The Old-Age Crisis—Actuarial Opportunities: The 1996 Bowles Symposium* Chapter 8, 72–79. Georgia State University, Atlanta.
- Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun. 2006. “Are Americans Saving “Optimally” for Retirement?” *Journal of Political Economy*, 114(4): 607–643.
- Scholz, John Karl, and Ananth Seshadri. 2012. “Health and Wealth In a Lifecycle Model.” University of Wisconsin-Madison Working Paper.
- Sheshinski, Eytan. 2007. *The Economic Theory of Annuities*. Princeton, New Jersey: Princeton University Press.
- Siegel Bernard, Tara. 2013. “Fine Print and Red Tape in Some Long-Term Care Policies.” *New York Times*, <http://goo.gl/NhRiX>.
- Sinclair, Sven, and Kent Smetters. 2004. “Health Shocks and the Demand for Annuities.” Congressional Budget Office Technical Paper 2004-09.
- Social Security Administration. 2005. “Social Security Disability Insurance Program Worker Experience.” *Actuarial Study No. 118*, Washington, D.C.
- Social Security Administration. 2011. “Social Security Disability Insurance Program Worker Experience.” *Actuarial Study No. 122*, Washington, D.C.
- Social Security Administration. 2012a. “Annual Statistical Supplement to the Social Security Bulletin, 2011.” Washington, DC.
- Social Security Administration. 2012b. “Disabled Worker Beneficiaries.” <http://goo.gl/YcM70>.
- Social Security Administration. 2013a. “Actuarial Life Table, 2009.” *Annual Statistical Supplement*, <http://goo.gl/FyspHb>.
- Social Security Administration. 2013b. “Estimated Size of the Population in the Social Security Area.” *Annual Statistical Supplement*, <http://goo.gl/MR8vjp>.
- Social Security and Medicare Boards of Trustees. 2012. “Status of the Social Security and Medicare Programs: A Summary of the 2012 Annual Reports.” <http://goo.gl/xXbuVL>.
- Standards and Poors. 2014. “2013 Annual U.S. Corporate Default Study and Rating Transitions.” <https://goo.gl/M7Upi0>.
- Starc, Amanda. 2012. “Insurer Pricing and Consumer Welfare: Evidence from Medigap.” The Wharton School Working Paper.

- Thaler, Richard H. 1991. *Some Empirical Evidence on Dynamic Inconsistency*. New York: Russell Sage Foundation.
- The Motley Fool. 2013. "Annuity Fees and Expenses." <http://goo.gl/F9xvK>.
- Townley, Peter G. C., and Robin W. Boadway. 1988. "Social Security and the Failure of Annuity Markets." *Journal of Public Economics*, 35(1): 75–96.
- Turra, Cassio M., and Olivia S. Mitchell. 2008. "The Impact of Health Status and Out-of-Pocket Medical Expenditures on Annuity Valuation." In *Recalibrating Retirement Spending and Saving*, ed. John Ameriks and Olivia S. Mitchell, Chapter 10, 227–250. Oxford University Press.
- U.S. Census Bureau. 2011a. "The Older Population: 2010." *2010 Census Briefs*, Washington, D.C.
- U.S. Census Bureau. 2011b. "Statistical Abstract of the United States: 2012." Washington, D.C.
- U.S. Census Bureau. 2012a. "Household Income Inequality within U.S. Counties: 2006–2010." Washington, D.C.
- U.S. Census Bureau. 2012b. "Income, Poverty, and Health Insurance Coverage in the United States: 2011." Washington, D.C.
- Warner, John T., and Saul Pleeter. 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review*, 91(1): 33–53.
- Warshawsky, Mark, Brenda Spillman, and Christopher Murtaugh. 2002. "Integrating Life Annuities and Long-Term Care Insurance: Theory, Evidence, Practice, and Policy." In *Innovations in Retirement Financing*, ed. Zvi Bodie, Brett Hammond and Olivia Mitchell, 198–221. University of Pennsylvania Press, Philadelphia, PA.
- Yaari, Menahem E. 1965. "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer." *Review of Economic Studies*, 32(2): 137–150.
- Zhang, Harold H. 1997. "Endogenous Borrowing Constraints with Incomplete Markets." *Journal of Finance*, 52(5): 2187–2209.

# FOR ONLINE PUBLICATION

## APPENDICES

### Appendix A: Proofs

#### Proposition 1

By equation (6),  $\rho_j(h) > r$  for all values of  $h$  provided that  $s_j(h) < 1$  (some people die).

#### Proposition 2

Inserting equation (1) into equation (3) and rearranging:

$$\begin{aligned}\rho_j(h'|h) &= \frac{1 + \pi_{j+1}(h')}{\frac{s_j(h)}{(1+r)} \cdot (1 + \sum_{h'} P(h'|h) \pi_{j+1}(h'))} - 1 \\ &= \frac{1 + \pi_{j+1}(h')}{\frac{s_j(h)}{(1+r)} \cdot (1 + E_H(\pi_{j+1}(h')))} - 1\end{aligned}$$

Because  $|H| > 1$  then  $\pi_{j+1}(\inf(H)) < E_H(\pi_{j+1}(h))$ . It is easy, therefore, to construct examples where  $\rho_j(h'|h) < r$ , thereby violating statewise dominance. Consider, for example, a set  $H$  with the elements  $h$  and  $h'$ , where  $s_j(h) \rightarrow 1$  and  $s_{j+1}(h') \rightarrow 0$  (and, hence,  $\pi_{j+1}(h') \rightarrow 0$ ). Then, we can further refine  $H$  so that  $E_H(\pi_{j+1}(h'))$  is sufficiently large, producing  $\rho_j(h'|h) < r$ , because  $E_H(\pi_{j+1}(h')) \rightarrow \infty$  implies  $\rho_j(h'|h) \rightarrow -1$ .

#### Proposition 3

The *expected* annuity return for a survivor to age  $j+1$  is equal to

$$\begin{aligned}E[\rho_j(h'|h)] &= \frac{1 + \sum_{h'} P(h'|h) \pi_{j+1}(h')}{\pi_j(h)} - 1 \\ &= \frac{(1+r)\pi_j(h)}{s_j(h)} - 1 \\ &= \frac{(1+r)}{s_j(h)} - 1 \\ &> r\end{aligned}$$

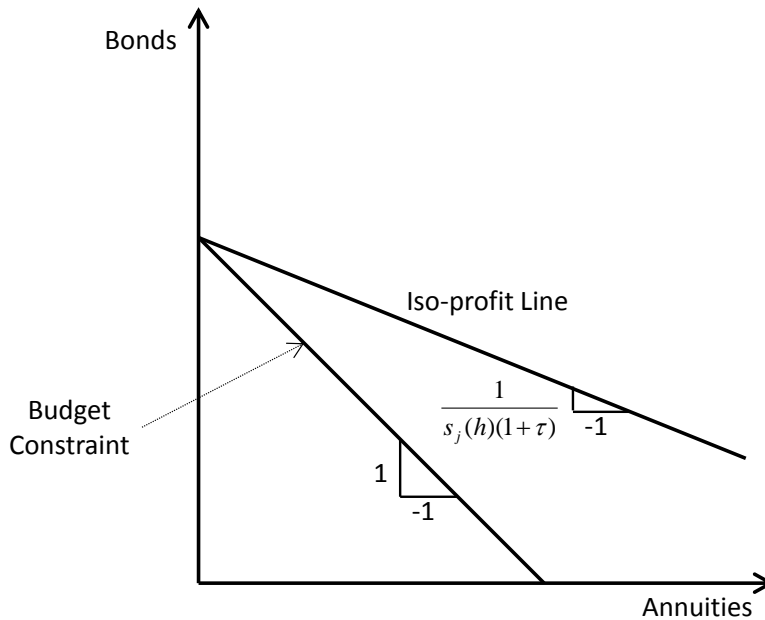
if  $s_j(h) < 1$ .

#### Proposition 4

See the example given in next subsection, which contradicts a claim of generic second-order stochastic dominance.



Figure 12: Optimal Annuitization in the Yaari Model with Transaction Costs



Explanation: Assumes that the transaction cost exceeds the value of the mortality credit.

## Appendix B: Robustness of Yaari’s Result to Additional Market Imperfections

### Transaction Costs, Moral Hazard, Social Security, Household Insurance and Uncertain Income

A couple of other market frictions can also rotate the Iso-profit Line. The most obvious one is *transaction costs*. Figure 12 shows the impact from adding a proportional transaction cost  $\tau$  that reduces the mortality credit, rotating the Iso-profit Line downward. In fact, if the differential transaction cost of annuities relative to bonds is so large that it actually exceeds the size of the mortality credit, then a risk-neutral agent will hold only bonds, as shown in Figure 12, where the Iso-profit Line now intersects the Budget Constraint at the 100% bond corner. In fact, annuitization is knife-edge (100% or 0%) in the Yaari model. *Moral hazard* could also rotate the Iso-profit Line if agents invest in living longer after annuitization. However, moral hazard cannot exist without annuitization; its corresponding Iso-profit Line must still intersect the budget constraint at the 100% annuity corner.

In fact, most commonly cited market frictions do not rotate the Iso-profit Line at all, thereby having no effect. Although *social security* crowds out some personal saving, the asset–annuity slope tradeoff for the remaining saving is unchanged. *Insurance within marriage* can reduce the level of precautionary saving, but it does not eliminate the statewise dominance of annuities for remaining saving. *Uncertain income* and *uncertain expenses*—whether correlated or not with *deterministic* changes in mortality probabilities—also have no impact on optimal annuitization.

## Appendix C: Robustness of Section 3

Although allowing for stochastic survival probabilities breaks the standard full annuitization result, allowing for a richer set of contracts could increase annuitization rates. We now consider a few.

### Shorter Contracts

In the three-period model, the annuity contract purchased at age  $j$  lasts until death or age  $j+2$ , whichever occurs first. Suppose, however, that we replace the two-period annuity contract with a sequence of one-period contracts, the first one issued at age  $j$  and the second issued at age  $j+1$ . There is no valuation risk with a one-period contract (formally,  $\pi_{j+1} = 0$  in equation (1)), and so the annuity return is simply equal to the bond yield plus any mortality credit, as in the original Yaari model. Annuities would again statewise dominate bonds.

Of course, from a welfare perspective, the value of the annuity diminishes with a shorter contract in the presence of reclassification risk. In the extreme case, with very short contracts approaching zero holding length, annuities provide no value because agents would simply rebalance right before they die. A mortality credit could not then be offered in a competitive equilibrium.

But we are more focused on annuity demand. Suppose agents also receive updates about their survival probabilities (and can die) at even a higher frequency than a single period. For example, the annuity contract might last for just one year, but the agent can receive information every six months. Then annuities will no longer dominate. Indeed, one can interpret our three periods as representing an interval of length  $\kappa$  in total time, with each period representing time length  $\frac{\kappa}{3}$ . Annuities will not dominate even as  $\kappa \rightarrow 0$  if information innovations occur at even higher frequency.

### A Richer Space of Mortality-Linked Contracts

Suppose now that households could also purchase additional mortality-linked contracts that make positive or negative payments based on changes in their individual health. Naturally, we will not consider an entire set of Arrow–Debreu securities; more rigid contracts like annuities exist precisely because a full set of Arrow–Debreu securities are not available. (In other words, a security that has any resemblance to a traditional-looking annuity would be spanned by existing securities in a full Arrow–Debreu economy.) Instead, we ask, what is the minimum type of mortality-linked contract that, when combined with an annuity, would restore annuities to their statewise (or even second-order) position of dominance?

For *patient* households, full insurance against all other shocks would restore full annuitization when there is no asymmetric information. For *impatient* households, recall the imperfect annuitization can happen even without correlated costs. In this case, additional payments would also need to be made to offset the pure annuity valuation risk, which is a non-observable cost in the standard sense. Such a security would need to be fairly rich in design and be a function of characteristics of previous and current health states and age (in order to capture duration).

### Hybrid and “Designer” Annuities

Thus far, we have considered a “life annuity” in the traditional sense, as a contract that pays a constant amount in each state contingent on survival, as in the original Yaari model.<sup>53</sup> Most of the annuity literature has focused on such a contract, which is our focus as well. It is straightforward, however, to con-

---

<sup>53</sup>Because we have no inflation in our model, we could also interpret our annuity payments as being indexed.

struct a “hybrid annuity” with bond-like features—specifically, one that includes some *non-contingent* payments—that will at least weakly dominate a simple bond. By subsuming both annuity and bond types of contracts, this hybrid annuity can never do worse than either a bond or a standard annuity, purely by construction.<sup>54</sup> Moreover, for impatient households facing no other risks, one could also create a “designer annuity” that makes contingent payments that decrease in real value with age, based on the agent’s own rate of time preference.<sup>55</sup> Finally, the demand for annuities could be altered if people could purchase an *option* contract that gave them the right to buy an annuity at a future date.<sup>56</sup>

## Appendix D: Discretization of State Space

Total wealth at age  $j$ ,  $A_j$ , is represented as one of 101 points of the wealth grid,  $A_{jk}$ ,  $k = 0, 1, \dots, 100$ . We fix point  $A_{j0} = 0$ ;  $A_{j100}$  equals the assumed maximum wealth, and the value of  $A_{jk}$  increases with  $k$ . For best interpolation during optimization and evaluation, the spacing between adjacent grid points is tighter at the low end of the wealth distribution, geometrically increasing values at intermediate to high wealth. Because most people’s wealth increases during the early part of life, the maximum wealth  $A_{j100}$  does not have to be the same for all ages; we also allow the grid to be expanded during the computation if the maximum wealth is actually reached by a positive measure of agents.

When the optimal policy (consumption, bond saving, and annuity saving) is computed for an agent at the node  $(A, \eta, h, j)$ , where the indices represent wealth, productivity, health, and age, respectively, the wealth  $A_{j+1}$  in the next period (age  $j + 1$ ) is allowed to take any positive value, rather than be limited to the values of the grid points. The value function  $V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j + 1)$  corresponding to that wealth is determined by interpolation between the two grid points bracketing it, for the given final productivity and health state  $(\eta_{j+1}, h_{j+1})$  and age  $j + 1$ . To reduce the potential for non-convexities induced by limited liability (i.e., Medicaid payments that ensure positive consumption), we set the minimum level of consumption sufficiently small to produce a monotone value function in wealth, thereby avoiding the artificial incentive to take on additional risk as wealth approaches zero. Still, to be extra careful, at each state within the household’s recursive problem, we execute a globally stable direct search optimization method numerous times across a wide range of different starting tuples along an appropriate mesh.

The number of nodes in the full dynamic-programming tree is  $(J - 20) \times m \times n \times (k^{max} + 1)$ , where  $(J - 20)$  is the age span between the minimum and maximum ages,  $m$  is the number of health states,  $n$

---

<sup>54</sup>Consider, for example, the case “Low Patience” ( $\beta \rightarrow 0$ ) considered earlier. A “hybrid annuity” that paid 0.75 at ages  $j + 1$  and  $j + 2$ , *not contingent* on actual survival, would allow the agent to consume 1.5 in both Good and Bad health states at age  $j + 1$ . The non-contingency of the payments allows even an agent in the Bad state to borrow at the zero risk-free rate against the payment that will be made at age  $j + 2$ , even though he or she does not survive until then. (If payments were *contingent* on survival, then the agent could never borrow in the Bad health state because the mortality-adjusted interest rate would be infinite.) The “hybrid annuity” would perfectly smooth consumption, as a bond does, by providing a non-contingent stream of payments. More generally, a “hybrid annuity” could reproduce any combination of bonds and traditional annuities when  $0 < \beta < 1$ .

<sup>55</sup>In the example considered earlier ( $\beta \rightarrow 0$ ), an annuity that paid a decreasing amount equal to 1.5 at age  $j + 1$  and 0 at age  $j + 2$  would again tie with a bond return. This decreasing-pay annuity, however, is different from a nominal annuity that makes decreasing real payments over time. Still, in practice, because a hybrid annuity is challenging to design, annuities paying a fixed nominal amount could be preferred over inflation-indexed annuities.

<sup>56</sup>See, for example, Sheshinski (2007), who nicely demonstrates a welfare improvement from the introduction of this unspanned contract when annuity contracts cannot be easily rebalanced. Aside from welfare changes, the impact on the actual demand for annuities in the model herein with rebalancing is ambiguous because of the trade-off between pooling reclassification risk early in life versus the value of obtaining more information about future mortality risk that has correlated costs in our setting. Regardless, annuitization must necessarily be less than full in equilibrium in our setting.

the number of productivity states, and  $k^{max}$  is the highest index of the wealth grid. We use ages from 21 to 120, so  $(J - 20) = 100$ ; as defined above,  $k^{max} = 100$ , and, as discussed in the paper,  $m = 3$  and  $n = 8$ . Therefore, we have about  $(J - 20) \times m \times n \times (k^{max} + 1) = 100 \times 3 \times 8 \times 101 = 242,400$  optimization problems for a single “partial equilibrium pass” of the household problem within the Gauss-Seidel routine, with each optimization problem computed up to 10 times with different starting values along a mesh. Obtaining a general equilibrium solution then typically requires 20 to 30 passes at the household problem. When the measure of agents is computed for the purposes of calculating aggregate quantities of capital and labor, a value from the continuum must be apportioned to the nearest two grid points. To preserve expected utility and the total measure, the weights given to the two points are chosen inversely proportional to the distance to them.

## Appendix E: Euler Equation Errors

Equation (11) can be rewritten more compactly as

$$V_j(A_j, \eta_j, h_j, j) = \max_{c_j, \alpha_j} \{u(c_j) + \beta s_j(h_j) \mathbb{E}_j[V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j+1)]\}$$

subject to the same budget constraints shown in the text. Assuming an interior solution the first order condition for consumption and ignoring intentional bequests ( $\xi = 0$ ) to simplify the exposition, implies that

$$\frac{\partial u(c_j)}{\partial c_j} = \beta s_j(h_j) \mathbb{E} \left[ \frac{\partial V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j+1)}{\partial A_{j+1}} R(\alpha_j, h_j, h_{j+1}) \right], \quad (12)$$

According to the Envelope Theorem the partial derivative with respect to  $A_j$  is

$$\begin{aligned} \frac{\partial V_j(A_j, \eta_j, h_j, j)}{\partial A_j} &= \frac{\partial u(c_j)}{\partial A_j} + \beta s_j(h_j) \mathbb{E} \left[ \frac{\partial V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j+1)}{\partial A_j} \right], \\ \frac{\partial V_j(A_j, \eta_j, h_j, j)}{\partial A_j} &= \beta s_j(h_j) \mathbb{E} \left[ \frac{V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j+1)}{\partial A_{j+1}} \frac{\partial A_{j+1}}{\partial A_j} \right], \\ \frac{\partial V_j(A_j, \eta_j, h_j, j)}{\partial A_j} &= \beta s_j(h_j) \mathbb{E} \left[ \frac{\partial V_{j+1}(A_{j+1}, \eta_{j+1}, h_{j+1}, j+1)}{\partial A_{j+1}} R(\alpha_j, h_j, h_{j+1}) \right], \end{aligned} \quad (13)$$

Noting that the right hand side of equations (12) and (13) are the same, we can rewrite

$$\frac{\partial u(c_j)}{\partial c_j} = \frac{\partial V_j(A_j, \eta_j, h_j, j)}{\partial A_j}$$

This allows us to rewrite equation (13) as

$$\frac{\partial u(c_j)}{\partial c_j} = \beta s_j(h_j) \mathbb{E} \left[ \frac{\partial u(c_{j+1})}{\partial c_{j+1}} R(\alpha_j, h_j, h_{j+1}) \right],$$

or

$$u_{c_j} = \beta s_j(h_j) \mathbb{E} [u_{c_{j+1}}(c_{j+1}) R(\alpha_j, h_j, h_{j+1})].$$

Solving for consumption, we get

$$c_j = u_{c_j}^{-1} \{ \beta s_j(h_j) \mathbb{E} [u_{c_{j+1}}(c_{j+1}) R(\alpha_j, h_j, h_{j+1})] \}$$

We now define the Euler Equation Error  $\varepsilon$  as

$$c_j(1 + \varepsilon) = u_{c_j}^{-1} \{ \beta s_j(h_j) \mathbb{E} [u_{c_{j+1}}(c_{j+1}) R(\alpha_j, h_j, h_{j+1})] \}$$

or

$$\varepsilon = \frac{u_{c_j}^{-1} \{ \beta s(h_j, j) \mathbb{E} [u_{c_{j+1}}(c_{j+1}) R_t(\alpha_j, h_j, h_{j+1})] \} - c_j}{c_j}$$

Generally, the acceptable range of errors is  $\log_{10}(\varepsilon) < -3$ . The Euler equation errors for people that are constrained—either because they live hand-to-mouth, or because they can annuitize only a positive fraction of their wealth—is typically larger than  $-3$ . The errors for unconstrained people typically range from around  $-3$  to less than  $-7$ .

## Appendix F: Estimating the Complementarity between Consumption and Health

To estimate the parameter  $\lambda$  used in Section 5.2.6 we use the Health and Retirement Study (HRS) and assign each observation to one of our health states based on their reported number of IADL and ADL impairments.<sup>57</sup> We then calculate the average number of chronic conditions for each of our three health states, which is what Table 15 shows. Based on the Finkelstein et al.'s (2013) finding that a 0.63 increase in the number of chronic conditions (a one standard deviation increase) is associated with a 10% to 25% decrease in the marginal utility of consumption, we calculate how the marginal utility of consumption would change as a result of transitions from healthy ( $h_1$ ) to impaired ( $h_2$ ), and from impaired ( $h_2$ ) to sick ( $h_3$ ). For example, when people in the HRS transition from  $h_1$  to  $h_2$ , the number of their chronic conditions increases by 1.04, at the average. That increase is larger by a factor of  $1.04/0.63 = 1.65$  than the increase of the number of chronic conditions the estimate by Finkelstein et al. is based on. Hence, we scale Finkelstein et al.'s reported decrease in the marginal utility of consumption up by a factor of 1.65, so that the appropriate range for our model is a decrease from 16% to 41% as shown in Table 16. Based on those estimates, and the functional form of our utility function, we back out the associated value of  $\lambda$  with a range from  $-0.26$  to  $-0.77$ . We also calculate the mid point of the estimates by taking the average of the lower and upper bounds.

## Appendix G: Additional Factors That Reduce Annuity Demand

### Management Fees

Yearly management fees for a typical annuity range from 0.80% to 2.0% of underlying assets, not including any initial commission charges (up to 10% of the base) or surrender fees (around 7% in the first year, declining by 1% per year thereafter).<sup>58</sup> In contrast, bond funds typically cost between 0.10% of assets (for an index of large firms) and 0.90% (for more specialized bonds, such as emerging markets). A differential management fee effectively reduces the mortality credit received from annuitization. We

<sup>57</sup>Recall that our health states for retirees are defined as in Robinson (1996): Healthy state  $h_1$  only includes people without impairments;  $h_2$  includes people who either have only IADL impairments, or who have no more than 2 ADL impairments;  $h_3$  includes people who have more than 2 ADL impairments or those who have some ADLs impaired and cognitive impairments).

<sup>58</sup>See The Motley Fool (2013) and CNN Money (2013).

Table 15: Number of Chronic Diseases in the HRS Data

Health State	Obs.	Mean	Std Dev.
1	141,779	1.74	1.56
2	28,415	2.78	1.78
3	10,336	3.29	1.90
Total	180,530	1.96	1.67

Source: HRS data for 1993-2012; Ages 50+. Definition of health states are based on the number of functional limitations in Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) as discussed in the main text and Robinson (1996).

assume a differential management fee of 1% and ignore commissions and surrender fees. Along the intensive margin, only 12% of wealth is annuitized. Of that, 27% of retiree wealth is annuitized and 4% of non-retiree wealth. Along the extensive margin, 38% of retirees hold a positive level of annuities but only 4% of non-retirees hold any annuities. These results are summarized in Table 17.

## Bequest Motives

### Without Management Fees.

Without an intentional bequest motive ( $\xi = 0$ ), all bequests are accidental and equal to about 2.6% of GDP in our baseline model with no management fees. Empirically, however, a ratio of aggregate bequests to GDP in the range from 2.0% to 4.0% per year is certainly reasonable (Gale and Scholz, 1994; Auerbach et al., 1995; Hendricks, 2002). We therefore consider the introduction of intentional bequests ( $\xi > 0$ ) and target a 3.3% bequest–GDP ratio, the point estimate of Auerbach et al. (1995). Management fees are initially set to zero. As shown in Table 17, now only 21% of wealth is annuitized (a decrease from the 38% shown in Section 5.1.3) and only 23% of households hold a positive level of annuities (a decrease from 26%).

### With Management Fees.

We also ran simulations that combined the same level of altruism  $\xi$  that produced a 3.3% bequest–income ratio with a 1.0% management fee. That combination increases the bequest–income ratio to 3.8%, at the upper bound of a reasonable range. In our baseline model, the amount of wealth annuitized in the economy dropped to 7%, with only 9% of households holding any annuities.

### With Management Fees and Uneven Bequests.

Empirically, only about 40% of the incidence of bequests are actually received as inheritances (Hendricks, 2002; Gale and Scholz, 1994). Some of the previous estimates of the bequest–income ratio do not clearly distinguish between *bequests* and *inheritances*. Therefore, we also ran simulations where only the bequests of the top 40% of income earners (as indicated by their wage at retirement) are received by younger higher-income earners. The other 60% is simply “thrown away” (for example, burial expenses). We target an *inheritance*–GDP ratio of about 2.7%, which produces an implied *bequest*–GDP ratio of around 5.2%. To be sure, this bequest–GDP ratio might be viewed on the high side. However,

Table 16: Implied Changes in the Marginal Utility of Consumption

Change in Number of Chronic Conditions	$\Delta$ in $u_c$			
	Lower Bound	Mid Point	Upper Bound	
<i>Finkelstein et al. (2013) estimate</i>				
0.63	-10%	-18%	-25%	
<i>Implied estimates based on results from Table 15</i>				
$h_1 \rightarrow h_2$	1.04	-16%	-29%	-41%
$h_2 \rightarrow h_3$	0.51	-8%	-14%	-20%
Implied $\lambda$	-0.26	-0.49	-0.77	

*Explanation:* Shows how the range of parameter values for  $\lambda$  are chosen such that the change in the marginal utility of consumption  $u_c$  when moving between model health states is consistent with estimates by Finkelstein, Luttmer and Notowidigdo (2013) that are based on the number of chronic conditions.

this value is actually close to the “lower bound” estimated by Gale and Scholz (1994) for households in the Survey of Consumer Finances. In particular, summing intended transfers, college expenses paid by parents, and accidental bequests, they estimate a ratio of annual flow of transfers to *capital* equal to 1.7%, which implies an annual flow to income ratio of about 4.7% at a capital–output ratio of 2.8. Along with a 1.0% differential management fee, only 4% of wealth is now annuitized in our model and only 7% of households hold any annuities (Table 17).

## Counterparty Risk

Beshears et al. (2012) report that the fear of default by annuity providers was another important factor discouraging survey responders from annuitizing. As the authors note, State-level guarantee funds do exist. Still, it is not unreasonable that many people might not fully trust or understand these guarantees. The guarantees are provided as a matter of policy and are not enforceable contracts. Indeed, unlike the FDIC, none of the state guarantees are even prefunded (except for NY, which historically has carried a small reserve). Instead, most states tax the policy premiums of the remaining insurers in order to fund shortfalls, with taxes typically capped at 2 - 3% of the premium. However, since annuity manufacturing is very concentrated, ex-post assessments might fail to recover enough funds if a large insurer collapses. Moreover, not all annuities (or all parts of a given annuity contract) are necessarily covered.<sup>59</sup> Quite reasonably, states do not want to invest resources ex-ante to determine which features will be covered for all contracts sold. Instead, states typically make the determination (and produce estimates of the actuarial value of the covered features) only after failure. Hence, it is reasonable that even people who are informed about the guarantee<sup>60</sup> might be concerned about the joint event of an insurer default and a

<sup>59</sup>A plain vanilla fixed annuity would almost certainly be covered, but those are less common.

<sup>60</sup>While insurers and their exclusive agents are not allowed to advertise the guarantee, independent financial advisers and broker-dealers do inform their clients about the guarantees and even layer annuities across providers since the guarantee is

Table 17: Changing Management Fees and Bequest Motives

Management fees	Bequest / GDP ratio	Annuitized fraction of wealth			Fraction of households with annuities			
		Total	Retirees only	Non-retirees only	Total	Retirees only	Non-retirees only	Fully Annuitized
<i>Without Intentional Bequests</i>								
0.00%	2.6%	38%	37%	38%	26%	52%	21%	9%
0.25%	2.8%	23%	35%	17%	18%	49%	11%	6%
0.75%	3.1%	13%	30%	5%	11%	41%	4%	5%
1.00%	3.3%	12%	27%	4%	10%	38%	4%	5%
1.20%	3.4%	11%	24%	4%	9%	35%	3%	4%
<i>With Intentional Bequests</i>								
0.00%	3.3%	21%	25%	19%	23%	49%	17%	6%
1.00%	3.8%	7%	15%	3%	9%	32%	3%	4%
1.00%	5.2%	4%	6%	2%	7%	20%	3%	3%

*Explanation:* Fraction of wealth annuitized and fraction of households with any annuities, for the entire population, retirees, and non-retirees at different levels of management fees and bequest motives.

*Assumptions:* Annuities are required to be non-negative. Social Security exists. Long-term care costs are 1.2 times average annual earnings. The capital–output ratio is set to 2.8 by varying the subjective discount rate.

state not paying, especially for an event that might happen decades after the purchase. Of course, people might also just misperceive the true default risk.

To examine the impact of counterparty risk, we started with our more optimistic economy (2.8 capital-output ratio) and examine what would happen if the insurer defaulted and a state paid nothing at a rate of 1 per 1,000 years (0.1% per year), implying a 15-year default rate of around 1.5%. A.M. Best Co. (2014) (Exhibit 2) estimates a 15-year “impairment” rate equal to 3.65 for “A++/A+” rated insurers, 6.53% for “A/A-”, 13.58% for “B++/B+”, increasing sharply for lower rated insurers.<sup>61</sup> However, “impairment” includes regulatory action before a default, and so these figures might be upward biased. Standards and Poors (2014) calculates a 0.49% weighted-average annual default rate for all U.S. insurers, across all ratings, between 1981 and 2013, for a 15-year average exceeding 7%. Our implied default rate, therefore, does not seem unreasonable. However, we are being a bit brutal here by assuming zero recovery value.<sup>62</sup>

Our results are reported in Table 18, where we set management fees and intentional bequests to zero. Previous models without correlated risks indicate that default risk has very little impact on annuitization (see, for example, the careful analysis by Lopes and Michaelides, 2007). Table 18 verifies that result by showing that counterparty risk affects annuitization only marginally in the Yaari version of our model without correlated risk. However, the effect of counterparty risk on annuitization is considerably larger in our model with correlated medical cost risk. Intuitively, despite the fact that the risks of a negative

subject to limits on the firm level rather than at the policyholder level.

<sup>61</sup>Using data from annuityadvantage.com, we estimate that about 43% of multi-year guaranteed deferred annuity contracts are sold by insurers with an A.M. Best Rating of A+ or higher. However, we don’t have volume information per contract.

<sup>62</sup>Alternatively, our assumptions could also be interpreted as a *perceived* risk of default. Under that interpretation, our assumptions are likely to be very conservative relative to the survey results reported in Beshears et al. (2012), where the fear of default plays an important role among consumers. However, we are not relying on this interpretation since our analysis is intended to be normative rather than positive.



Table 18: Annuitization with Potential for Default: Intensive and Extensive Margins

	Intensive Margin	Extensive Margin	Fully Annuitized
<i>Baseline Calibration, Default Risk of 1 in 1,000 Years</i>			
Non-retirees	24%	18%	3%
Retirees	22%	43%	1%
Total	23%	23%	2%
<i>Yaari Calibration, Default Risk of 1 in 1,000 Years</i>			
Non-retirees	95%	95%	73%
Retirees	98%	93%	83%
Total	95%	94%	76%

*Explanation:* Shows the amount of annuitization when the model allows for a chance of full default. A default rate of 0.1% corresponds to an insurer defaulting and the State not paying anything 1 time every 1,000 years, or a cumulative default probability over 30 years of 3 percent.

*Assumptions:* Annuities are required to be non-negative. Social Security exists. Long-term care costs are 1.2 times average annual earnings. The capital–output ratio is set to 2.8 by varying the subjective discount rate.

health shock and insurer default are uncorrelated, there is now a chance that an insurer default happens at the same time as a negative health shock occurs that dramatically lowers the annuity value, just when annuitants “really need the money” to pay for correlated uninsured costs. Even very small probabilities when interacted with very large marginal utility states can have material effects.

## Potential Future Extensions

We now consider three possible extensions that would likely decrease the demand for annuities even more. We tried to implement each of them but faced computational challenges or limited access to the household level of data that would allow for a clear model calibration. Therefore, we leave these extensions up to future research.

### Differential Transaction Costs.

Another possible extension would incorporate differential product transaction costs above the management fees considered earlier. Actual transaction fees for investing in bond funds are quite low, ranging from zero at vertically integrated broker—dealers such as Vanguard to small ticket charges at independent broker—dealers such as Schwab and Fidelity. In contrast, transaction costs for buying an insurance product such as an annuity are larger. In addition to the initial underwriting charge for determining a client’s risk profile, the presence of health shocks in our model means that rebalancing would require additional underwriting in order to reduce adverse selection. These factors should further reduce the level of annuitization. Incorporating such one-sided transaction costs into our model would be computationally very challenging and is left to future research.<sup>63</sup>

<sup>63</sup>In particular, shape preservation of the value function is not well defined in higher dimension. We spent a considerable amount of effort on various approximation methods without success, as deemed by the Euler errors.

### **More Worker Risk.**

Recall that workers in our model were assumed to always qualify for disability insurance to partly cover their lost wages as well as private insurance to fully cover their medical costs. As a result, the only risk that workers face from health shocks in our model is from the portion of their wages that is not covered by disability. In reality, workers face risk in the form of negative health shocks that reduce future wages without becoming disabled. Workers also face uninsured medical costs in the form of low coverage or copayments. We could not find the micro-level data that would allow us to map these additional risks along the key dimensions of our model; the available data appears to be too aggregated.

### **Asymmetric Information.**

Finally, recall that our simulations assume that policyholders do not hold superior information relative to insurers. As we showed earlier, while adverse selection reduces the mortality credit, it does not undermine the case for full annuitization in the Yaari model. Even an annuity with a smaller mortality credit statewise dominates bonds in the Yaari model, producing a corner solution at 100% annuitization. However, in the model herein with stochastic mortality probabilities and correlated costs, most households face an interior condition in their choice between annuities and bonds. As a result, any reduction in the mortality credit from asymmetric information would tend to reduce positive annuitization even more. That could result, for example, if the insurer does not want to incur the costs associated with medical underwriting. If short sales are allowed, then shorting by younger households could also be undermined if their subsequent opportunity to take a positive position is limited.