

NBER WORKING PAPER SERIES

OPTIMAL ANNUITIZATION WITH STOCHASTIC MORTALITY PROBABILITIES

Felix Reichling
Kent Smetters

Working Paper 19211
<http://www.nber.org/papers/w19211>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2013

This paper is based, in part, on Sven H. Sinclair and Kent Smetters, Health Shocks and the Demand for Annuities, Technical Paper 2004-09 (Congressional Budget Office, July 2004). The authors have benefited from discussions with Zvi Bodie, Michael Boskin, Jeffrey Brown, Daniel Gottlieb, Caroline Hoxby, Laurence Kotlikoff, Robert Merton, Rick Miller, Joshua Rauh, Dan Sacks, Tom Sargent, Florian Scheuer, Joseph Tomlinson, and seminar participants at the 2012 Brigham Young University Computational Public Economics Conference, the Congressional Budget Office (CBO), Boston University, and Stanford University. 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Felix Reichling and Kent Smetters. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Optimal Annuitization with Stochastic Mortality Probabilities
Felix Reichling and Kent Smetters
NBER Working Paper No. 19211
July 2013
JEL No. D01,D14,H31

ABSTRACT

The conventional wisdom dating back to 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. Annuities still help to hedge longevity risk, but they are now subject to valuation risk. Valuation risk is a powerful gateway mechanism for numerous frictions to reduce annuity demand, even without ad hoc "liquidity constraints." We find that most households should not annuitize any wealth. The optimal level of aggregate net annuity holdings is likely even negative.

Felix Reichling
Macroeconomic Analysis Division
Congressional Budget Office
United States Congress
Washington, DC 20515
felix.reichling@cbo.gov

Kent Smetters
University of Pennsylvania
SH-DH 1354
3620 Locust Walk
Philadelphia, PA 19104
and NBER
smetters@wharton.upenn.edu

1 Introduction

The classic paper of Yaari (1965) demonstrated that the demand for annuities should be so strong that life cycle 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 fairly uncommon.¹ Indeed, the apparent under-annuitization of households is commonly referred to as “the annuity puzzle” (see, for example, Modigliani, 1986; Ameriks et al., 2011). This puzzle is not just a theoretical curiosity. The value of the mortality credit can be very large later in life, thereby significantly increasing the return to saving. Furthermore, when an investor’s preferences exhibit prudence, annuities also reduce the need for precautionary savings, thereby 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 also face health costs (Sinclair and Smetters, 2004) and liquidity constraints after annuitization (Bodie, 2003; Davidoff, Brown and Diamond, 2005; Turra and Mitchell, 2008). 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, if the deterministic survival probabilities of the Yaari model are taken seriously, then 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 traditional model does not allow for a negative [positive] health shock during the life cycle to suddenly decrease [increase] an investor’s likelihood of survival.

In this paper, we largely adopt the Yaari framework but allow for the mortality probabilities themselves to be stochastic. This modification 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 health shocks (Palumbo, 1999; De Nardi, French and Jones, 2010). It is natural to investigate how such shocks also affect decisions to annuitize. We show that allowing for stochastic mortality probabilities can have a material impact on the optimal level of annuitization.

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 valuation risk (or principal or resale risk), much like a long-term bond. Specifically, the present value of the annuity

¹Sales 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).

stream falls after a negative shock to health that reduces a household's life expectancy. Some previous theoretical research has shown that agents will want to pool even this source of risk—known as “re-classification risk”—by purchasing the lifetime annuity early in the life cycle rather than later (see, for example, the original work by Brugiavini, 1993). Hence, the presence of stochastic mortality probabilities alone would not appear to undermine the case for full annuitization. This result may explain why the annuity literature has largely ignored the role of stochastic mortality probabilities.

However, the previous literature that discussed stochastic mortality probabilities implicitly made two important assumptions: (a) negative health shocks that decrease the value of the annuity are not correlated with any additional costs, and (b) agents are sufficiently patient—that is, they do not significantly discount future utility. Empirically, however, negative health shocks often produce correlated costs such as a loss in income and/or uninsured medical costs. Moreover, households are not necessarily fully patient. 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.

We show that relaxing either assumption can lead to imperfect annuitization. The introduction of correlated costs reduces the optimal level of annuitization below 100% even if households do not face any ad hoc “liquidity constraints.” Intuitively, even if agents can borrow against the present value of their annuity stream, the value of the annuity falls after a negative health shock. That decline in value happens precisely when the household needs resources to pay for the correlated-cost shocks in order to avoid an increase in its marginal utility. We will call this effect the *correlated-cost channel*.

We further show that this channel also serves 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 adjust premiums for selection in practice.

We then demonstrate that even the presence of correlated costs can be dropped altogether if households are sufficiently *impatient*. In this case, the presence of valuation risk alone, even without correlated costs, implies that the annuity will fail to smooth consumption across those states that are actually most valued by investors. We will call this effect the *impatience channel*.

We then present simulation evidence using a multi-period life cycle model that demonstrates the power of the correlated-cost channel itself in reducing annuity demand. (In our baseline simulations, households are sufficiently patient, thereby shutting down the “impatience channel”.) Our model 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 that the correlated-cost channel has on reducing annuity demand.

Our simulation results demonstrate that the correlated-cost channel is quite effective at reducing the optimal level of annuitization. Annuitization is 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 63% of households should *not annuitize any wealth*, even with no transaction costs, ad hoc “liquidity constraints,” or bequest motives. In contrast, the Yaari model predicts 0%.

In fact, the *net aggregate* annuitization is actually *negative* under the most unconstrained version of our simulation model, where households are also allowed to sell (short) annuities. 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 short annuities even with *no* bequests motive. Under plausible calibrations, the negative

demand for annuities by younger households outweighs the positive demand by older households, producing a net negative aggregate demand across the measure of households.

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-cost channel risk. We find that younger savers should instead do the opposite: actually 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.² The difference in the value of these short-long offsetting trades produces a net profit to the household that can then be used to pay for any correlated income loss and/or uninsured expenses.

In sum, relative to the conventional 100% annuitization result, more detailed optimal life cycle annuitization patterns emerge from our model where: (a) most households do not annuitize any wealth; (b) positive [negative] annuitization by non-wealthy households is largely concentrated in those households that can earn a large [small] mortality credit relative to costly valuation risk; and (c) positive annuitization is more likely in wealthy households where the costs that are correlated with a decline in health are small relative to their assets.

This paper does not intend to explain all of the stylized facts surrounding annuities, including the design of annuity contracts (Gottlieb, 2012). Some recent interesting research has examined 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 in the Yaari framework is relaxed. Of course, as with most academic research, this paper is not intended by itself to change the practice of financial advisory, but rather to help lay the foundation for additional study.

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.³ To investigate this issue further, we introduce some additional real-world factors into the model, including reasonable levels of asset management fees and bequest motives. Our results show that almost nine out of ten households in our model do not positively annuitize any wealth. We further argue that the 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, including among households with no dependents, rather than the lack of positive 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,

²The mechanics are discussed in more detail in subsection 2.2.3. 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.

³For 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.

the presence of stochastic mortality probabilities implies that annuities deliver a larger expected return (from the mortality credit) along with *more* risk (from valuation risk). A greater level of risk aversion, therefore, typically *reduces* annuitization in our model.

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 presents concludes.

2 Three-Period Model

Consider an individual age j in 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 state h at time j . State h is drawn from a countable set H with a cardinality exceeding 1. We can interpret these elements as "health states," although, more generally, they can represent anything that affects survival probabilities. The cardinality assumption ensures that there is more than one such state, and so we model the Markov transitional probability between states 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 π_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 (that is, with fair pricing) where insurers can perfectly pool idiosyncratic mortality risk, the premium paid at age j must equal the *actuarial present value* of the payment of 1 received from the annuity in each of the future two periods:

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

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. 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:

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

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:

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

In other words, the off-diagonal elements of the corresponding Markov transition matrix $[P(h'|h)]$ are zero, recognizing the absence of health reclassification risk. 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:

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

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

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

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 is 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, annuities statewise dominate bonds for any initial health state at age j .*

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

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 strong and 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 “Insurance Line” in Figure 1 shows the tradeoff between bonds and fairly priced annuities that would be offered by a competitive annuity provider. The slope of the insurance 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, $\$1 \cdot (1 + \rho_j) = \frac{\$1}{s_j(h)} \cdot (1 + r)$.

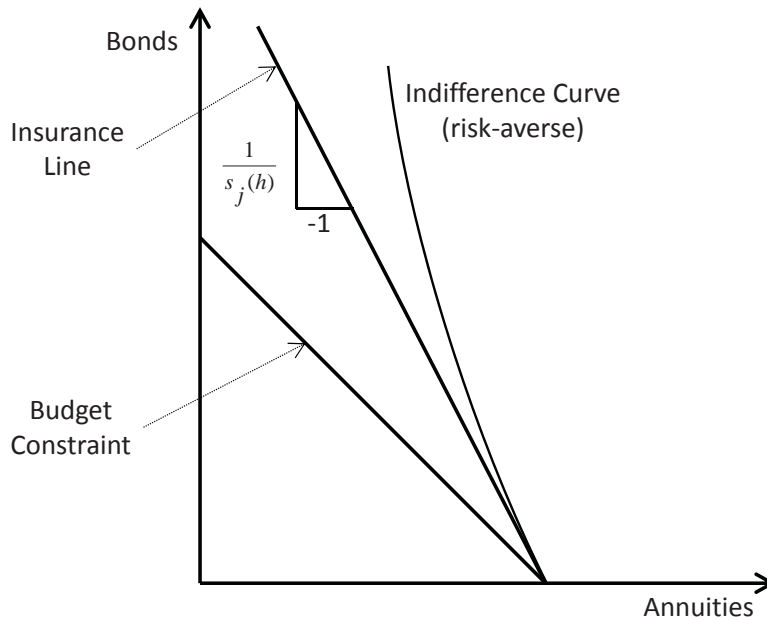
The Insurance Line is also the indifference curve between bonds and annuities for a *risk-neutral* investor.⁴ Specifically, a risk-neutral investor would be willing to give up \$1 in annuity investment if she could trade it for $\frac{\$1}{s_j(h)} > \1 worth of bonds, because both investments would have 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 insurance 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.⁵

⁴Incidentally, 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.

⁵At 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 Insurance 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 Insurance Lines.

Figure 1: Optimal Annuitization in the Yaari Model



2.2.1 Example: Adverse Selection

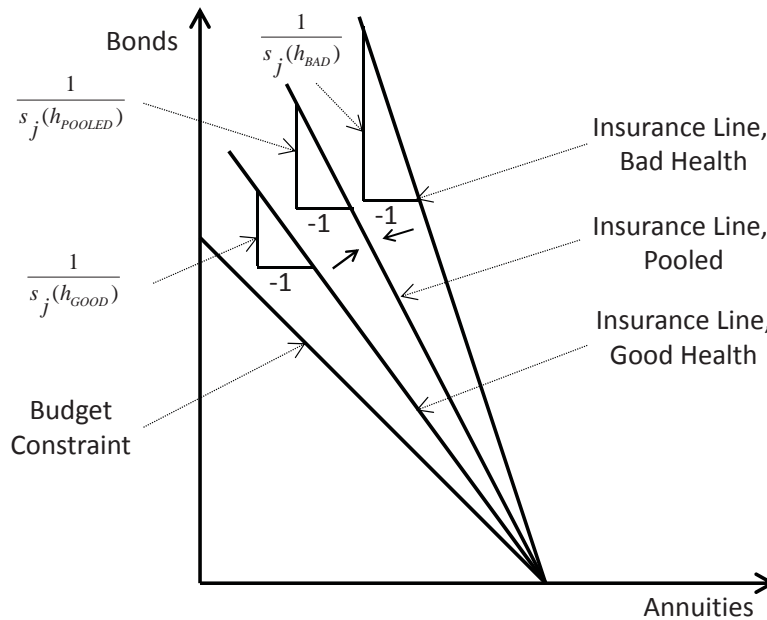
The “corner optimality” of the Yaari model is hard to break. Figure 2 illustrates how the corner optimality is robust to the presence of adverse selection. Suppose that health h at age j can take on two states: Bad health, h_B , and Good health, h_G , where, naturally, the probability of survival is 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 Insurance 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 Insurance 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 Insurance 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 Insurance Line for Good health shows the tradeoff for people with Good health. Naturally, the Insurance 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 Insurance Line in Figure 2.⁶ The effect of this pooling is that households with Bad health experience a loss in annuity return, indicated by a downward rotation in their Insurance Line. Household with Good health experience a gain in annuity return, indicated by an upward rotation in their Insurance Line. But notice that full annuitization for both types still occurs, despite the cross-subsidy, because each Insurance

⁶Mathematically, 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 insurance line is also a competitive equilibrium, provided that there is no other annuity provider that can better identify the individual risk types.

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



Line 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 Other Market Imperfections

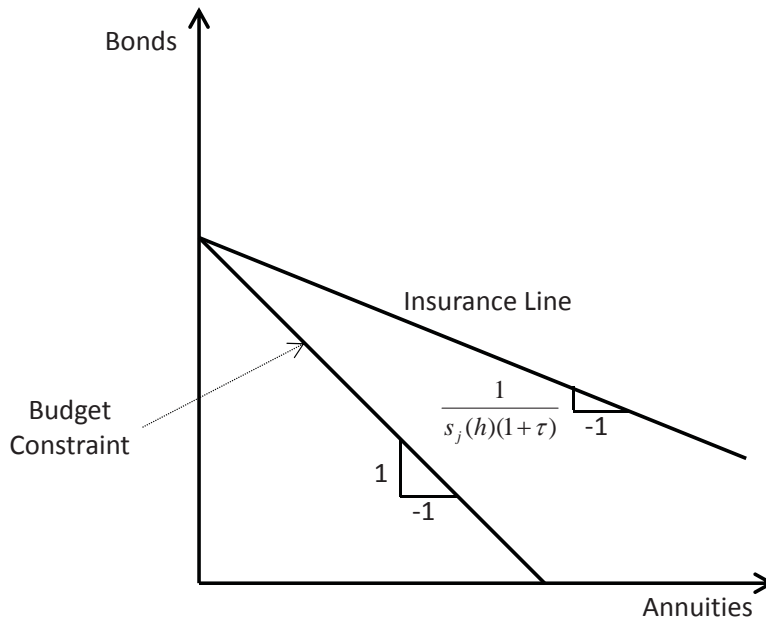
A couple of other market frictions can also rotate the insurance line. The most obvious one is *transaction costs*. Figure 3 shows the impact from adding a proportional transaction cost τ that reduces the mortality credit, rotating the insurance line downward. In fact, if the differential transaction 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 3, where the Insurance 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 insurance line if agents invest in living longer after annuitization. However, moral hazard cannot exist without annuitization; its corresponding insurance line must still intersect the budget constraint at the 100% annuity corner.

In fact, most commonly cited market frictions do not rotate the insurance 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.

2.2.3 “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

Figure 3: Optimal Annuitization in the Yaari Model with Transaction Costs



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

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.⁷ 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 actually imposing a very different sort of 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 a much stronger assumption than a standard borrowing constraint noted above. 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.⁸ 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

⁷Most 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).

⁸Even 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.

life insurance contract, and then borrow the present value of the life insurance contract, π_{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 is not altogether clear. Recall that the standard annuity puzzle is the small size of the *primary* 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.⁹ Moreover, as just noted above, a person can reverse an annuity simply by purchasing 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.¹⁰

Moreover, 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 $h' \neq h$. In other words, the off-diagonal elements of the Markov transition matrix $[P(h'|h)]$ are allowed to be positive.

3.1 Stochastic Rankings

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

⁹We could not find any aggregate industry information on the annuity secondary market. However, firms such as J.G. Wentworth actively advertise to purchase retirement annuities, as do other firms, suggesting a strong supply. Moreover, many life settlement firms, which normally buy life insurance policies, will also purchase annuities.

¹⁰See 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. They found very small differences between the originally estimated (curtate) life expectancies and the actual death dates.

Proposition 2. *With stochastic survival probabilities ($p(h'|h) > 0$), annuities do not generically statewise dominate bonds.*

Proof. 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 $\rho_j(h'|h) \rightarrow -1$. \square

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, when survival probabilities are stochastic, the annuity contract has valuation risk (or principal risk) similar to a long-dated bond. If participants can directly sell their contracts in a secondary market, then we can also interpret this valuation risk as “resale” risk.

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. It is still possible, however, that annuities could dominate bonds for more specific types of preferences, including expected utility maximizers. Indeed, annuities will be strictly preferred by one class of expected utility maximizers—namely, *risk-neutral* consumers.

Proposition 3. *The expected return to annuities exceeds bonds if the chance of mortality is positive.*

Proof. 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 $s_j(h) < 1$. \square

Intuitively, risk-neutral agents care only about the greater expected return derived from the mortality credit and ignore valuation risk. Any *expected* change in an agent’s survival status has already been

priced by competitive insurers into the initial annuity premium. Any residual uncertainty that stems from the future survival outlook taking a value different from the expected value today has no impact on the expected return to the annuity.

However, annuities do not necessarily dominate bonds for the more restricted class of *risk-averse* expected utility maximizers.¹¹

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

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

3.2 Examples

We now demonstrate Propositions 2 through 4 with a series of simple examples that build on each other. We first demonstrate how the statewise dominance of annuities fails before turning to the failure of second-order stochastic dominance.

3.2.1 Failure of Statewise Dominance

Continuing with our three-period setting, consider an agent at age j with current health state h . As an example, we make several simplifying assumptions.¹² Specifically, set the bond net return r (and, hence, discount rate) to 0, in order to simplify the present value calculations. Also, at each age:

- 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 survive to collect the \$1 annuity payment at age $j + 1$. This assumption eliminates the potential to earn any positive mortality credit between ages j and $j + 1$, which is not our focus.
- 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: $s_{j+1}(h_B) = 0$. In words, a person who realizes Good health at age $j + 1$ will with certainty also live in period $j + 2$, but a person realizing a Bad state will with certainty not live in period $j + 2$.

The payoffs for the annuity are summarized in Figure 4. By equation (1), the competitive premium paid at age j for an annuity is equal to $\pi_j(h) = \$1.0 + 0.5 \cdot \1.0 , or \$1.5. This amount is simply equal to the \$1 annuity payment that is received with certainty at age $j + 1$ plus the expected value of the \$1 annuity payment made at age $j + 2$, which is paid 50% of the time to people who realize Good health at age $j + 1$.

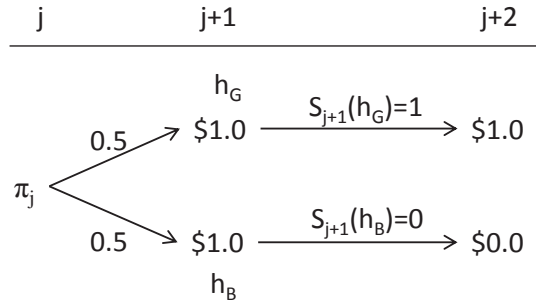
Suppose that an agent, therefore, is thinking about whether to invest \$1.5 in the annuity or bond. Consider two cases:

■ 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 equal to $\frac{2}{1.5} - 1 > 0$, thereby beating bonds

¹¹Specifically, gamble A second-order stochastically dominates gamble B if and only if $E_A u(x) \geq E_B u(x)$ for any nondecreasing, concave utility function $u(x)$.

¹²The main advantage of using simple examples is that we can isolate the key drivers of our results. However, none of the key insights in this section hinge on the simplifications we made.

Figure 4: Annuity Payoff in Simple Example



which, recall, yield a 0% net return. Specifically, the annuity (before any payouts) is worth \$2 at age $j + 1$, equal to the \$1 that will be paid at $j + 1$ plus the present value (at a zero discount rate) of the \$1 in annuity payment that will be 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 at age $j + 1$. So this household is better off *ex post* with the annuity.

■ Case 2 (“Bad” health): The household realizes the Bad health state h_B at age $j + 1$. Then, the realized net return to annuitization is equal to $\frac{1}{1.5} - 1 < 0$, thereby under-performing bonds. Specifically, the annuity will pay only \$1 in total at age $j + 1$, the last year of life. Had this household instead invested \$1.5 into bonds, then it would have had \$1.5 at age $j + 1$. So, it is worse off *ex post* with the annuity.

It follows that annuities fail to statewise dominate bonds in the presence of valuation risk because annuities fail to dominant bonds *ex post* across some states. Intuitively, the competitively priced annuity contract at age j was calculated based on *expected* survival outcomes at age j . Survival realizations below expectation at age $j + 1$ 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 the case of 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, however, could still be the dominant security for *risk-averse expected utility* maximizing agents whose preferences fully weigh risky gambles from an *ex ante* position—that is, at age j . Indeed, having some *ex post* losers is the cost of providing *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 very robust. Suppose that our agent at age j with health h is endowed with \$1.5 and consumes only in ages $j + 1$ and $j + 2$.¹³ The agent has fairly standard time-separable *conditional* expected utility preferences over consumption equal to

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

where the period felicity function $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ takes the constant relative risk aversion (CRRA) form, σ

¹³This 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 $j + 1$, thereby allowing us to capture the motive to pool reclassification risk.

is the level of risk aversion, and β is the weight placed on future utility. As will be evident below, our analysis holds for any risk-averse function. However, the CRRA assumption allows us to report a few numerical examples as well.

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 $s_{j+1}(h_{j+1} = h_B) = 0$ (that is, Bad health people do not live to age $j+2$).

As we now show, the agent's subjective rate of time preference plays an important role in the choice between annuities and bonds in the presence of valuation risk. So we consider two cases: with highly patient agents and then with agents with low patience. Within the case of patient agents, we consider two subcases: with and without correlated costs. These subcases produce different decisions to annuitize.

High Patience ($\beta = 1$). Suppose that agents are fully patient, weighting future utility equal to current utility: $\beta = 1$. Consider two subcases, one without correlated costs and one 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 the \$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 the following:

■ **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 $c_{j+1} = 1.5$.¹⁴

■ **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 $c_{j+1} = 1.0$.¹⁵

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).¹⁶

¹⁴Specifically, 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.

¹⁵Recall that the annuity in this example pays \$1 in each period that an agent survives.

¹⁶This 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.

II) With Correlated Costs. Now consider the introduction of uninsured shocks 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 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. With smaller correlated costs, partial annuitization would emerge. Simulation analysis is presented later using a more realistic calibration.

The presence of stochastic probabilities, therefore, opens up the possibility for other market frictions, such as uninsured costs, to materially reduce the demand for annuities. We now show that these correlated costs have an impact on annuitization, which we are referring to as the *correlated-cost channel*.

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 $j + 1$. Hence, there is no future annuity income to borrow against at age $j + 1$. Of course, we can easily generalize this result by allowing for a small but positive chance of survival after age $j + 1$ after the Bad health state is revealed. In this case, the annuity has some continuation value at age $j + 1$, which could then be borrowed against. But it still might not provide enough resources to avoid states with low consumption at age $j + 1$ if the household fully annuitizes. Appendix A discusses how this imperfect annuitization result can extend to variations in the design of the annuity contract.

In sum, additional costs that are correlated with a reduction in survival can lead to incomplete annuitization on their own. In contrast, with *deterministic* survival probabilities as in the Yaari model, the correlated-cost channel is absent since the lack of a health update eliminates the channel for a correlated reduction in the present value of the annuity.

Low Patience ($\beta \rightarrow 0$). Now consider the opposite extreme, where agents are very impatient: $\beta \rightarrow 0$.¹⁷ We now show that incomplete annuitization can emerge even without the presence of additional correlated uninsured costs.

The conditional consumption streams (without correlated costs) are now as follows:

■ **Bond:** If Good health is realized at age $j + 1$, then $c_{j+1} \rightarrow 1.5$ and $c_{j+2} \rightarrow 0$; if Bad health is realized, then $c_{j+1} = 1.5$.

■ **Annuity:** If Good health is realized at age $j + 1$, then $c_{j+1} \rightarrow 2.0$ and $c_{j+2} \rightarrow 0$; if Bad health is realized, then $c_{j+1} = 1.0$.

For the bond investment, consumption at age $j + 1$ with Good health is 1.5, because a fully impatient agent consumes everything at age $j + 1$. If Bad health manifests, the agent obviously still consumes 1.5 regardless of the level of patience, because age $j + 1$ is the last year of life. For the annuity investment,

¹⁷Because discounting is geometric, the results do not rely upon time inconsistency issues.

consumption at age $j + 1$ in the Good state is 2.0, because the impatient agent borrows against the annuity payment of 1.0 that will be paid for sure at age $j + 2$. However, consumption at age $j + 1$ in the Bad state is now just 1.0, because the annuity does not make any payment at age $j + 2$.

Notice that the annuity now actually *increases* consumption variation across the two health states at age $j + 1$; in contrast, the bond investment perfectly smooths consumption across those allocations that are actually *valued* by the agent. Intuitively, the traditional annuity fails to properly shift consumption across *health states* in a way that is informed by the agent’s time preferences *within* each health state. The fair (competitive) annuity premium is based on *objective* survival probabilities and the market discount rate, which can result in a very different set of allocations across time within health states than subjectively desired. In other words, a traditional life annuity is not an optimal contract design for very impatient households. This *impatience channel*, therefore, reduces the optimal demand for annuities. This channel does not exist in the Yaari model, because the health state is fixed; the level of patience affects only the desired level of saving and not how to invest the saving.

Of course, the choice of $\beta = 0$ is an extreme case and is made for purposes of illustration. Now suppose that the value of β is small but still strictly positive. Hence, the marginal utility associated with consumption at age $j + 2$ must also be considered. In general, the highest expected utility is obtained when we equalize *the weighted marginal utilities* across different states, which is not equivalent to equalizing consumption when $\beta < 1$. Given the analysis above, the Bond can still be more effective at smoothing the weighted marginal utilities as well. As a specific example, suppose that $\beta = 0.10$.¹⁸ Also, set the level of risk aversion σ equal to 2. Then the $EU(\text{bonds}) = -0.91$ and the $EU(\text{annuities}) = -0.93$. Of course, if we were to change the value of σ then the curvature of the utility function becomes more important and could tilt the balance back toward the annuity.

3.3 A Gateway Mechanism

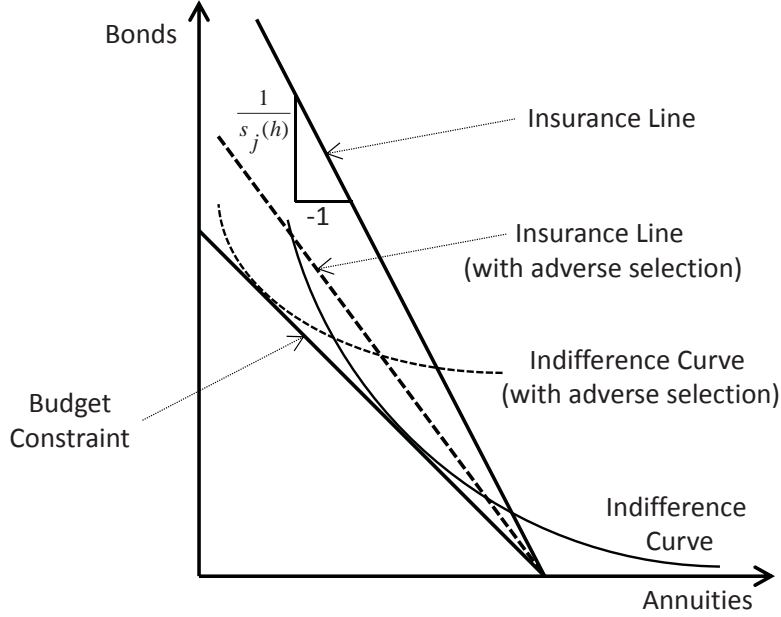
As just shown, stochastic survival probabilities provide a mechanism for additional frictions to reduce the optimal level of annuitization. Below, we provide simulation analysis using a richer setting. However, it is useful to see graphically how the presence of stochastic mortality probabilities can remove the sharpness of 100% “corner optimality” found in the Yaari model. We focus on the case of patient investors facing costs that are correlated with negative mortality shocks.

Figure 5 shows that the Indifference Curve between bonds and annuities can now take the more usual “convex-toward-the-origin” property of a standard risk-averse investor. Recall that, in the Yaari model, the Insurance Line represents the lower bound of the Indifference Curve. However, in the presence of valuation risk with correlated 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 Insurance 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 Insurance Line downward, as shown with the dotted Insurance Line in Figure 5. The Indifference Curve must then rotate along the budget constraint, also shown in Figure 5, in order to ensure that at least one region of the Indifference Curve has a flatter slope than the corresponding Indifference Curve. The net effect is a higher demand for bonds.

¹⁸If we interpret each period in our simple model as 30 years, then this small value is equivalent to an annual rate of about 0.92.

Figure 5: Optimal Annuitization with Health Shocks



4 Multi-Period Model

We now present a multiple-period model that provides the basis for simulation evidence presented in Section 5.

4.1 Individuals

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

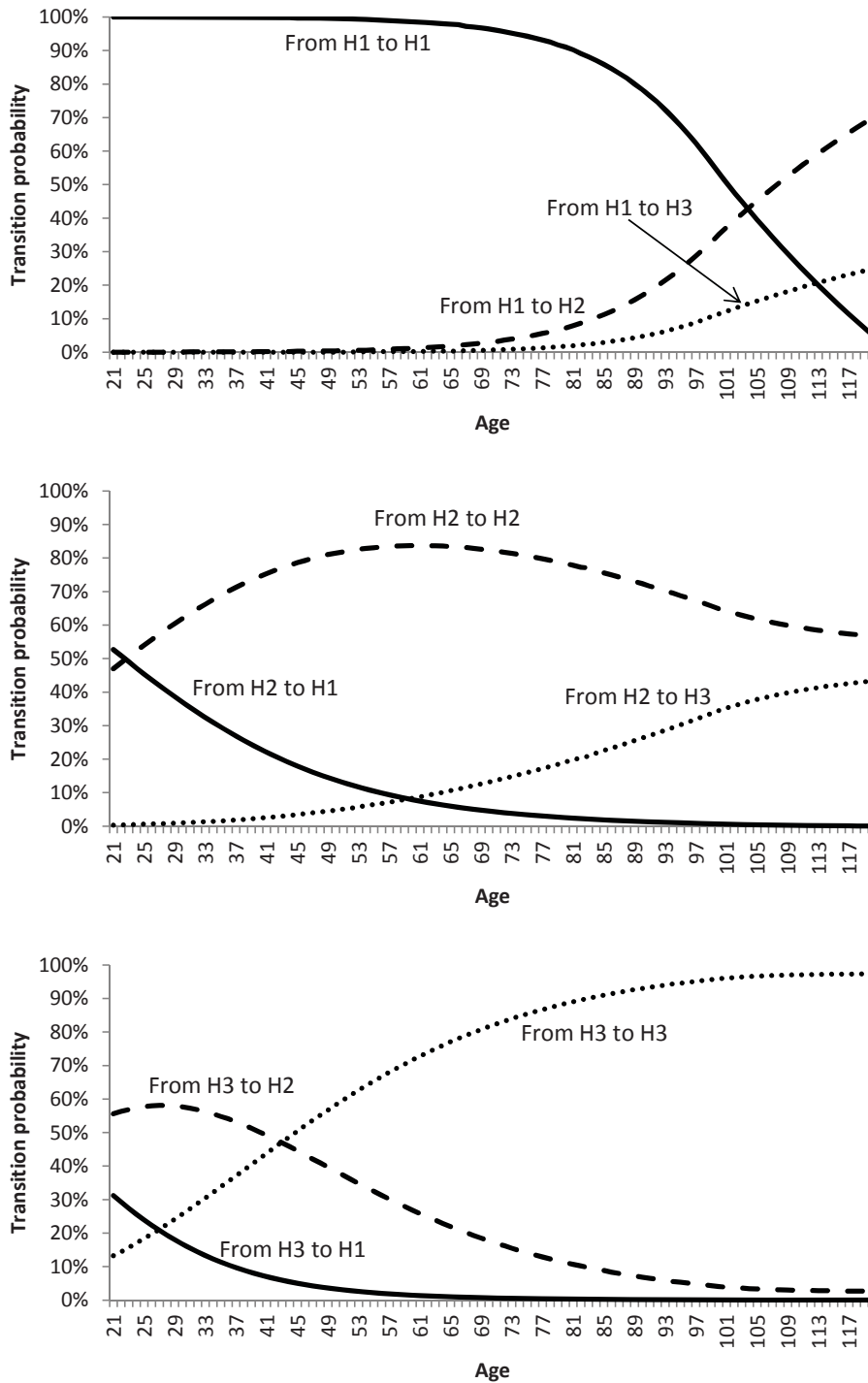
4.1.1 Health Transition Probabilities and Conditional Survival Probabilities

The health state h follows an M -state Markov process with an age-dependent transition matrix $P_{mn}(j)$; $m, n = 1, \dots, M$, where n is the previous health state and m is the next health state. For our purposes, three states ($M = 3$) suffice: healthy (h_1), disabled (h_2), and very sick (h_3). Most workers are healthy (state h_1) and able to work. A disabled worker (state h_2) is unable to work and receives disability benefits until retirement, as discussed in more detail below. This state, therefore, represents impaired health with higher mortality risk, but without actually being institutionalized and suffering the associated expenses. A very sick person (state h_3) is unable to work and faces institutionalized expenses, but receives disability payments until retirement and, potentially, additional transfers to cover long-term care expenses.

Transition and survival probabilities are based on the actuarial model of Robinson (1996).¹⁹ Transi-

¹⁹Robinson'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, h_2 corresponds to his states 2–4, and h_3 corresponds to his

Figure 6: Health Transition Probabilities



Source: Authors' calculations based on references cited in Section 4.1.1. Health state H1 corresponds to healthy, H2 to disabled, and H3 to very sick.

Table 1: Disability Rates: Data vs. Model

Age Cohort	U.S. Data	Model
21-24	0.3%	0.5%
25-29	0.9%	1.0%
30-34	1.6%	1.5%
35-39	2.2%	2.2%
40-44	3.2%	3.3%
45-49	4.8%	5.1%
50-54	7.1%	7.6%
55-59	10.3%	10.6%
60-64	13.8%	13.9%
Total	5.0%	4.2%

Source: Social Security Administration and authors' calculations.

tion probabilities between health states are shown in Figure 6. Robinson's estimates are widely regarded as the industry standard for older workers and retirees, but he under samples younger workers. We used smoothed interpolations of Social Security disability data to slightly adjusted his health state transition estimates for the working-age population, thereby producing disability rates by age that are fairly close to the empirical evidence (see Table 1).²⁰ Overall, the share of the working-age population that is disabled (that is, either health state h_2 or h_3) in our baseline model equals about 4.2%, which is close to the actual disability rate of 5.0% found in the Social Security Administration's disability data.²¹

Survival probabilities by age and health state are shown in Figure 7. The model's mean mortality rate, averaged across health states at each age, matches the U.S. Social Security Administration tables fairly closely.²²

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 discussed in more detail below. Bonds, therefore, constitute a safe investment in our model because the aggregate capital stock is deterministic. However, in any experiments that change the amount of capital, the value of the risk-free rate also changes, reflecting changes to the marginal product of capital.

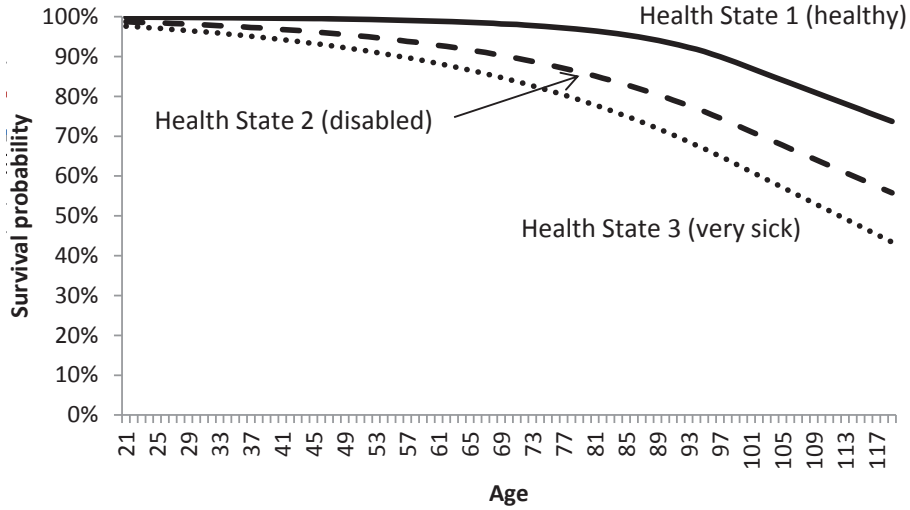
states 5–8, which he notes is consistent with additional medical expenditures. Future work could consider additional sick states beyond long-term care, provided that these states could be mapped to a specific set of chronic conditions and their associated uninsured costs. Currently, such data is not readily available.

²⁰Because SSA data do not indicate the exact health status of a disabled person, we conservatively made adjustments only to the transition probabilities between health state h_1 and health state h_2 , thereby leaving the change in mortality to a minimum.

²¹See Social Security Administration (2012b) and Social Security Administration (2011).

²²We also did some comparisons of model assumptions against their empirical counterparts within each health state. Because the very sick state h_3 corresponds most closely with long-term disability during working years and with a need for long-term care for retirees, the sources of mortality data comparisons included the RP-2000 disabled life mortality tables (Society of Actuaries, 2000a,b) for ages 21 to 65, as estimated by the Social Security Administration (2005).

Figure 7: Survival Probabilities



Source: Authors' calculations based on references cited in Section 4.1.1.

Annuities. Households can also invest in an annuity that pays \$1 per unit contingent on survival. Figure 8 shows the realized single-period net annuity return ρ as a function of age and health-state transition. This return includes two components: (a) for a *given* health state, the annuity earns the standard *mortality credit* and (b) after a *change* in health state, the annuity incurs a *valuation* repricing: either a depreciation if health worsens or an appreciation if health improves. The value of the mortality credit increases monotonically with the death probability over the life cycle. As a result, the annuity return increases over the life cycle *within* an unchanging health state (except for the last maximum possible year of life, at age 120, where a value depreciation is guaranteed). The valuation risk, however, creates level shifts in the return. Notice, for example, that movements from healthy state h_1 to the worsening health states h_2 and 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, as indicated earlier in Figure 6.

4.1.3 Income and Expenses

An individual's "cash on hand" X_j at the beginning of age j is equal to

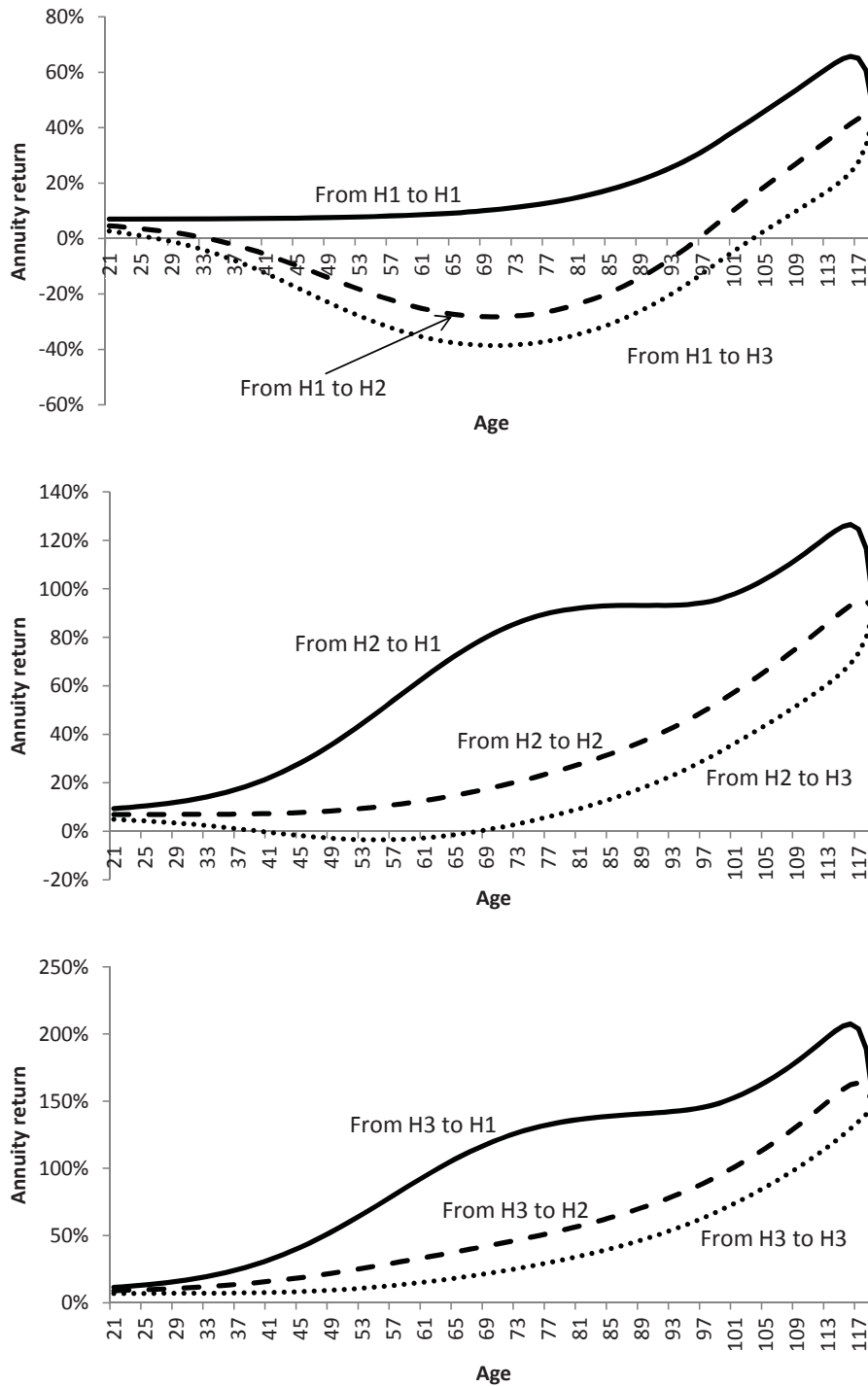
$$X_j = \varepsilon_j \eta_j I(h = h_1) w(1 - T) + B_j - L_j + Tr_j, \quad (9)$$

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

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

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

Figure 8: Annuity Returns by Age and Health State Transition



Explanation: The panels show the annuity return ρ 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 a disabled 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 ε_j and η 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 description below.

Bequests. The variable B 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 that are not covered by Medicare, equal to about 1.2 times the average wage, a value that is consistent with recent survey estimates conducted by Genworth Financial (2012) and MetLife (2010). This calibration likely understates the true cost of long-term care by assuming 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, long-term policies have tended to be 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).²³ 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. In its latest annual survey, Fidelity Benefits Consulting (2012) estimates that a representative couple retiring in 2012 will face almost \$230,000 in health care expenses that are not reimbursed by Medicare. These costs do not include nursing home 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 by 30% above costs, and less than one third of retirees have such policies (Starc, 2012). Even retirees with long-term care 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.

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 don't have a precise mapping between our health state Markov transition matrix and health costs unrelated to long-term care. However, we will consider different parameterizations of the financial loss L_j as part of our sensitivity analysis.

²³We thank Caroline Hoxby for this point.

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 conservatively assume that all disabled workers qualify for this income, and without a waiting period. In reality, the U.S. DI program requires a significant waiting period and 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). *Hence, the only correlated costs faced by workers from a negative health care shock in our model are from the portion of income that is not replaced by DI.*

Retirees receive Social Security benefits according 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.²⁴ Following the law, future Social Security benefits cannot be borrowed against. However, following a medical loss ($L > 0$), if cash on hand falls enough then Medicaid will pay the medical costs, thereby ensuring that the household never suffers from negative consumption. *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 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.

Overall, we view our modeling choices as likely conservative in the sense of avoiding states with very large marginal utilities after a negative mortality shock.

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 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:

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

where β is the rate of time preference, c_j is consumption at age j , σ is the risk aversion, A_j is bequeathable wealth at age j , D_j is an indicator that equals 1 in the year of death and 0 otherwise, ξ is a parameter that determines the strength of the bequest motive, and $J = 120$, the maximum age.

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

²⁴Hence, 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 the last working year the expected benefit conditional on the income earned in last year, whether for disability or retirement. 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.

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

subject to:

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

$$\alpha \leq 1$$

$$0 \leq c_j \leq A_j + X_j$$

where: A_j is beginning-of-period asset, after annuities and bonds are sold; the portfolio return is $R(\alpha_j, h_j, h_{j+1}) = \alpha \rho_j(h_j, h_{j+1}) + (1 - \alpha)r$, 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 rebalancing; r is the return to the riskless bonds; α_j is the share of investments made into annuities at age j ; X_j is the value of cash-on-hand shown earlier in equation (9). Here, $A \in \mathbf{R}_+$, $\eta \in \mathbf{D} = \{\eta_1, \eta_2, \dots, \eta_n\}$, $h \in \mathbf{H} = \{h_1, h_2, h_3\}$, $j \in \mathbf{J} = \{21, 22, \dots, J\}$, and the functions $\{V_j, a_j, b_j, c_j : \mathbf{S} \rightarrow \mathbf{R}_+\}_{j=21}^{120}$ 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 σ -algebra of \mathbf{R}_+ .

The budget constraints have the following interpretations. Bonds 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 the loan. However, annuities 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. Moreover, an individual's consumption c must always remain non-negative. Without health shocks, this 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. Nonetheless, we want to ensure that our utility preferences are consistent with a plausible capital–output ratio that is, in turn, derived from a reasonable technology specification. Moreover, having a firm production side of the economy allows us to recalibrate to the same observable economy when performing sensitivity analysis.

Output Y of the economy is determined by the constant returns-to-scale technology with a Cobb–Douglas production function $Y = \theta K^\lambda N^{1-\lambda} - \delta K$, where K denotes capital and N labor, in efficiency units, and δ is the rate of capital depreciation. 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 N . Macroeconomic variables are also calibrated consistent with Nishiyama and Smetters (2005): The capital share of output is $\lambda = 0.32$, the depreciation rate of physical capital is $\delta = 0.046$, and the capital-to-output ratio is 2.8. The rate of population growth is assumed to be a constant 1 percent,

roughly the value in the United States. The rate of productivity increase is also not critical and, for the sake of simplicity, is assumed to be unchanging. The capital-output ratio is set at 2.8 by varying the subjective discount rate β , producing a marginal product of capital of 6.8%.

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

The entire recursive household partial equilibrium dynamic 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 B and C for more details. Achieving small Euler errors required using optimization routines that were slower but have good global properties. Still, using FORTRAN 90, a machine compiled language that generally executes 50 to 500 times faster than MATLAB, the fixed point can usually be reached within two days on a fast desktop computer, although some of our experiments below took several days.

4.5 The Implied Population, Income, and Wealth Distributions

In our baseline model, annuities are required to be non-negative (no shorting), an assumption that we will relax as part of our sensitivity analysis. Moreover, all bequests (if any) are accidental ($\xi = 0$) and are distributed evenly to surviving households, an assumption that we also relax. We also initially assume there are no differential management fees on annuities relative to bonds. We now examine how the model calibrates to some observable aggregate distributions.

Population Distribution. The role of population demographics is less important in our model because we report our results mainly at the household level (age and health status). However, our claim below that the net aggregate annuity holding is negative depends, in part, on our model's relative population weights. The age structure of our model is fairly similar to 2010 Census data, as shown in Table 2.

Table 2: Age Structure of the Population, Data vs. Model

Age Cohort	2010 Census	Model
21-29	17%	21%
30-39	18%	21%
40-49	20%	19%
50-59	19%	16%
65+	18%	17%

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

Income Distribution. Our baseline model seems to calibrate fairly well to the observable data on income inequality. The income Gini coefficient (inclusive of wage income, Social Security, DI, and other benefits) is 0.45 in our model, compared with 0.47 in the data (U.S. Census Bureau, 2012a). Moreover, about 5.0% of workers in our model are above the payroll tax ceiling, compared with 5.4% in the data Social Security Administration, 2012a.

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 0.61, in contrast to the empirical estimate reported by Nishiyama (2002) of 0.75. Table 3 shows the share of wealth held by different wealth percentile groups for both the model and the corresponding empirical evidence from the Census. Notice that 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 19.6% of wealth (10.9% vs. 30.5%), which, in turn, persists throughout the "Top 40%" of the wealth distribution. We narrow the gap somewhat in our sensitivity analysis when we turn on intentional and unequal bequests, but we never eliminate it.

The Rich. Life cycle models similar to the model in our study 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 (Cagetti and De Nardi, 2006). We believe that any reasonable model of "entrepreneurial spirit," however, would only strengthen our key conclusion regarding low optimal annuitization. As shown later, a majority of households do not hold annuities. Instead, almost all annuities are held by wealthier households, which are mostly insulated from large shocks to their marginal utilities after negative changes in their mortality risk outlook. Because the rich in our model have no entrepreneurial motives, they are attracted mainly to the larger returns offered by the lifetime annuity wrapper. With a realistic entrepreneurial motive, however, they likely would not hold nearly as many annuities. For example, it would be quite challenging to design an annuity wrapper around an individual founder's privately held equity.

The Poor. The total *poverty rate* among *all workers* between the ages of 18 to 64 in our model is 4.2%, which is a bit below the Census value of 7.2% (U.S. Census Bureau, 2012b). Our underestimate is likely because the U.S. Census has a broader definition of "working" than our model's wage data have. However—and more important for our purposes—the *poverty rate among all disabled people* between the ages of 18 to 64 is 33.5% in our model, which is pretty close to the empirical counterpart of 28.8%, estimated by U.S. Census Bureau (2012b), although lower than the value of 50% estimated

Table 3: Wealth Distribution, Data vs. Model

	U.S. Data ⁽¹⁾	Model
Gini Coefficients		
Income ⁽²⁾	0.47	0.45
Wealth	0.75	0.61
Share of Wealth (%)		
Top 1%	30.5	10.9
Top 5%	53.9	29.2
Top 10%	64.9	42.2
Top 20%	77.2	60.3
Top 40%	90.4	85.7
Top 60%	96.9	96.5

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

by Congressional Budget Office (2012). 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, therefore, that we are not overestimating its impact and, if anything, our calibration is on the conservative side.

5 Simulation Results

We now present simulation evidence from the multi-period model, starting first with our baseline calibration discussed above, followed by some sensitivity analysis. 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, weighted by the share of households in each health state at that age.

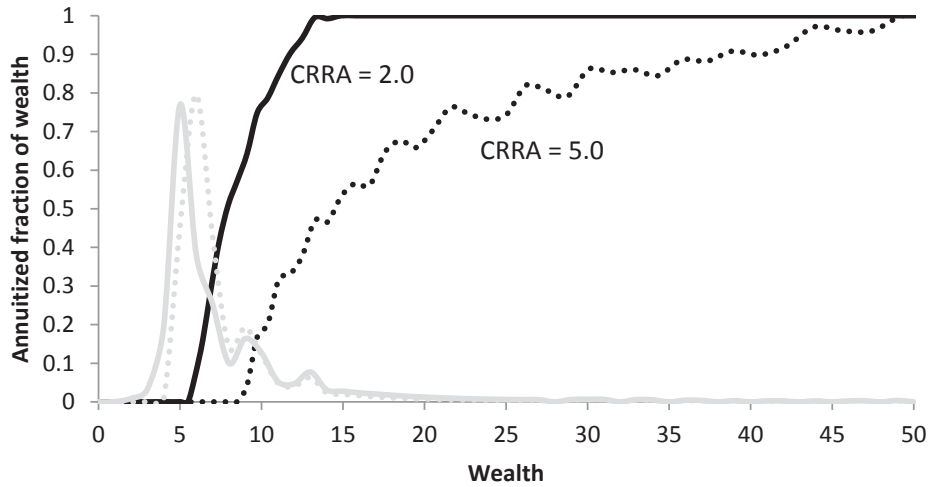
5.1 Baseline Model

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

5.1.1 Annuitization at Age 65

The darker lines in Figure 9 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 wage in the economy. Notice that the level of annuitization varies significantly by the level of risk aversion σ and wealth. For households with $\sigma = 2$, a relatively low level of risk aversion, annuities are not purchased at levels of wealth below around 6 times the national average annual earnings. For households with a higher level of risk aversion of $\sigma = 5$, annuities are not purchased at levels of wealth below 8 times average annual earnings. Notice, therefore, that a larger risk aversion

Figure 9: Annuitized Fraction of Wealth at Age 65 for a Healthy Person



Explanation: The black lines show the optimal fraction of wealth held in life annuity form by a healthy person ($h = 1$) age 65 with different coefficients of risk aversion, plotted as a function of beginning of period wealth. The unit of wealth is the average annual earnings. The gray lines with matching plot patterns represent the distribution of households with constant relative risk aversion (CRRA) of $\sigma=2$ [$\sigma=5$].

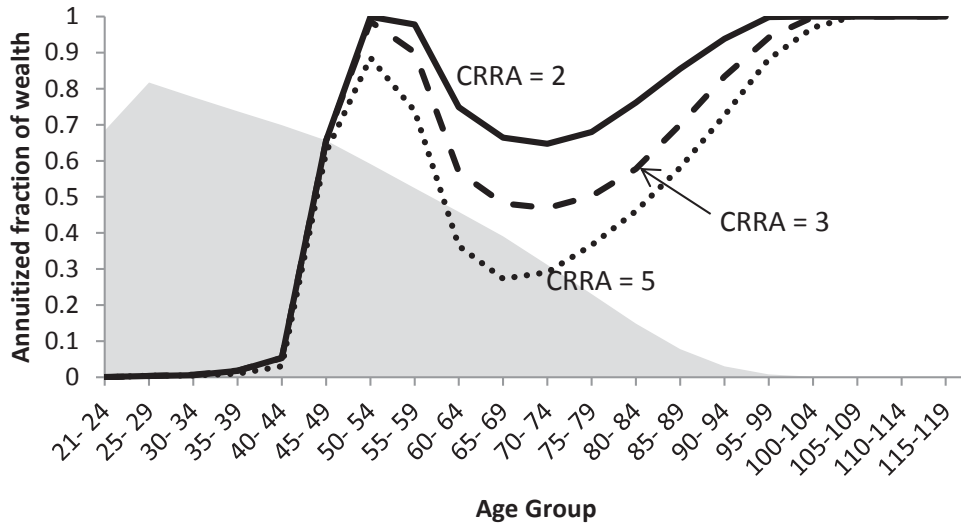
Assumptions: Annuities are required to be non-negative. Social Security exists. Long-term care costs are equal to 1.2 times average annual earnings. Asset management fees and bequest motives are absent. The capital–output ratio is set to 2.8 by varying the subjective discount rate.

σ decreases the demand for annuities, contrary to the standard Yaari model. This pattern is driven by the valuation risk in our model that produces the *correlated-cost channel*, which is absent in the Yaari model. To investigate, we ran the Yaari version of our model without health shocks and found that all age-65 households annuitized 100% of their assets at all levels of wealth for both $\sigma = 2$ and $\sigma = 5$. Moreover, when we ran our model with health transitions but assumed that long-term care costs were fully insured, all wealth at retirement was also fully annuitized, suggesting that the *impatience channel* is not driving these results.

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.) Hence, a wealthy retiree can “hold to maturity,” in much the same way that a long-term bond holder is less concerned with duration risk.

However, most retirees do not have a sufficient level of wealth that allows them to hold to maturity. The gray lines in Figure 9 show the percentage of age-65 households with the indicated level of wealth. Much of this distribution falls to left of the point of any positive levels of annuitization and very far to the left of the points of full annuitization. In fact, with $\sigma = 3.0$, a value between our low and high levels of risk aversion (not plotted, to reduce clutter), 53% of retirees (across all health states) at age 65 hold no annuities at all and only 11% fully annuitize. Recall that these results are not driven by ad hoc liquidity constraints; agents can fully rebalance their asset portfolio. Rather, these results are produced by the correlated-cost channel, where the increase in the uninsured health costs coincides with the fall in the remaining value of the annuity.

Figure 10: 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.

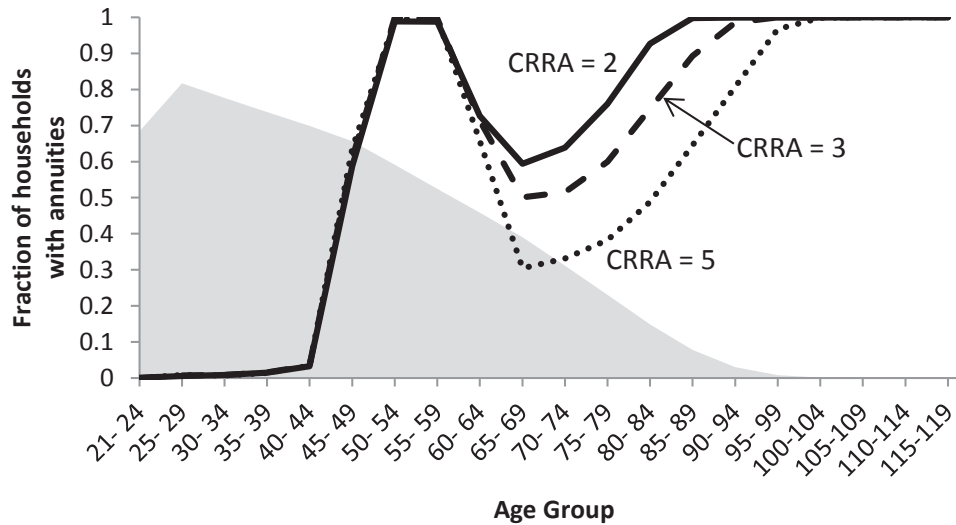
5.1.2 Annuitization across the Life cycle

Figure 10 shows the fraction of all wealth that is annuitized (the intensive margin) across the life cycle. Figure 11 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 independent of the level of risk aversion.

Across all ages, almost 57% of wealth is annuitized at $\sigma = 3.0$. However, only 37% of all households in the economy hold any positive level of annuities; the other 63% hold none. Moreover, only 24% of households fully annuitize. How can the majority 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 the costs associated with negative mortality shocks out of the annuity income stream itself.

Figures 10 and 11 show that annuitization, though, is not monotonic in age. Figure 8, which shows the realized annuity return ρ as a function of age and health-state transition, provides the intuition for this pattern. Recall that the annuity return includes two components: (a) the standard mortality credit, which monotonically increases over the life cycle as the probability of death increases for a *fixed* health state, and (b) the valuation risk associated with *changes* in the health state that alter the remaining value of the annuity. Consider first the 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 annuitizing. The chance of moving from good health to worse health may not be large in absolute terms, but it is *relative* larger than the value of the small mortality credit. Most younger households choose, therefore, not to annuitize, mostly out of fear that the value of the annuity might fall at the same time that they might suffer correlated costs (mainly lost wages). Now consider persons above age 95. They earn a large mortality credit even if they happen to be healthy (h_1); the credit is even larger if they

Figure 11: 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.

are not currently healthy (h_2 or h_3). A larger fraction of older people are also unhealthy initially, and so the valuation risk is even smaller. Older people, therefore, typically choose to annuitize their entire wealth.

5.2 Sensitivity Analysis

We now consider several changes to the baseline model that allow us to investigate the importance of our long-term care cost assumptions, the presence of Social Security, and the impact of allowing for a short position in annuities. In each case, full annuitization would still exist in the Yaari model.

5.2.1 Long-Term Care Costs

Table 4 investigates the importance of changing uninsured long-term care costs away from our baseline value of 1.2 times average annual earnings. Smaller values are consistent with some households (about one in ten) having long-term care insurance. Larger values are consistent with the fact that our baseline analysis ignored other forms of non-insured health care costs that are likely correlated with the health state; our baseline analysis also assumed that, at most, only one member of the household would use long-term care. Recall that, for *working* agents, we conservatively assume that all long-term care costs are paid for by private insurance. Hence, changes in long-term care costs mainly impact the annuitization decisions of retirees.

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 44% of households in the economy still want a positive level of annuitization even with no long-term care costs. In contrast, in the corresponding

Table 4: 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	55%	45%	37%	34%	31%	30%
1.50	61%	51%	41%	36%	34%	32%
1.20	67%	57%	45%	39%	37%	33%
0.60	75%	71%	61%	43%	42%	40%
0.00	75%	73%	69%	44%	44%	44%
	<i>Retirees Only</i>					
1.80	47%	32%	21%	56%	38%	27%
1.50	56%	40%	25%	64%	48%	31%
1.20	69%	51%	33%	74%	61%	39%
0.60	95%	88%	71%	91%	86%	74%
0.00	100%	100%	100%	91%	91%	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.

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.

Yaari model (not shown), all households fully annuitize in the presence of uninsured expenses because there are no ad hoc constraints on asset rebalancing.

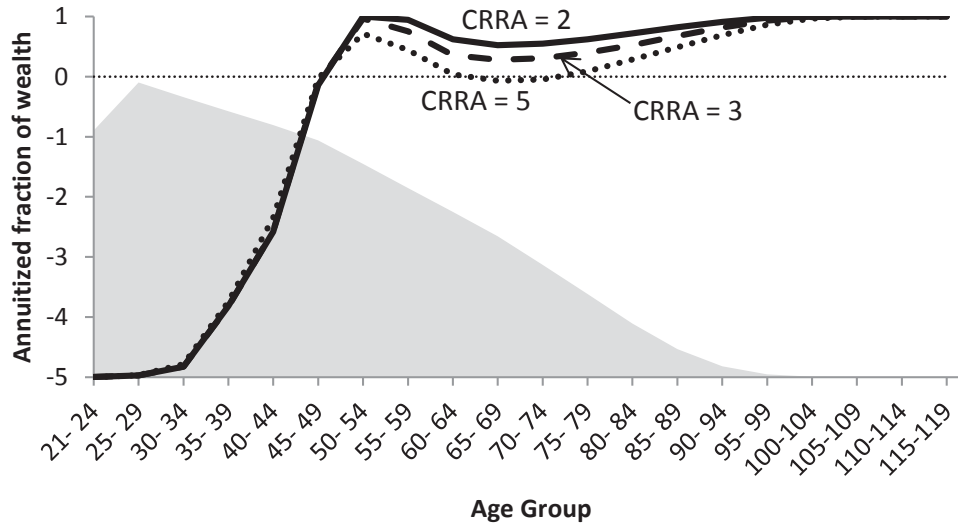
5.2.2 Social Security

In the Yaari model with deterministic survival probabilities, Social Security crowds out household saving, but it does not change the investment allocation toward annuities.

In our model with stochastic mortality probabilities, however, Social Security has more of a material impact. On one hand, Social Security is not perfectly substitutable for private annuities because Social Security cannot be rebalanced (it is illegal to borrow against future benefits). The more that a household is required to save into Social Security, the less after-tax income that is available for precautionary savings against correlated cost shocks. On the other hand, Social Security provides a valuable annuity that implicitly incorporates the population average mortality credit (through the payroll tax rate).

For the most part, these two effects offset each other in our model, although the first effect is larger. In particular, upon removing both the Social Security payroll tax and benefit at current factor prices, the amount of wealth annuitized at $\sigma = 3.0$ declines from 57% in our baseline model to 46%. The fraction of households with any annuity remains almost unchanged at 36%. The fraction of households that fully annuitize, however, drops from 24% to just 11%. Interestingly, these results suggest that Social Security has the potential to *crowd in* private annuitization, contrary to the conventional view.

Figure 12: 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. The value of $\sigma = 3.0$.

5.2.3 Allowing for Short Selling

Thus far, our simulation analysis has restricted annuity allocations to be non-negative. This 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. As discussed earlier, a negative annuity is equivalent to borrowing at the risk-free rate and purchasing life insurance to ensure repayment.

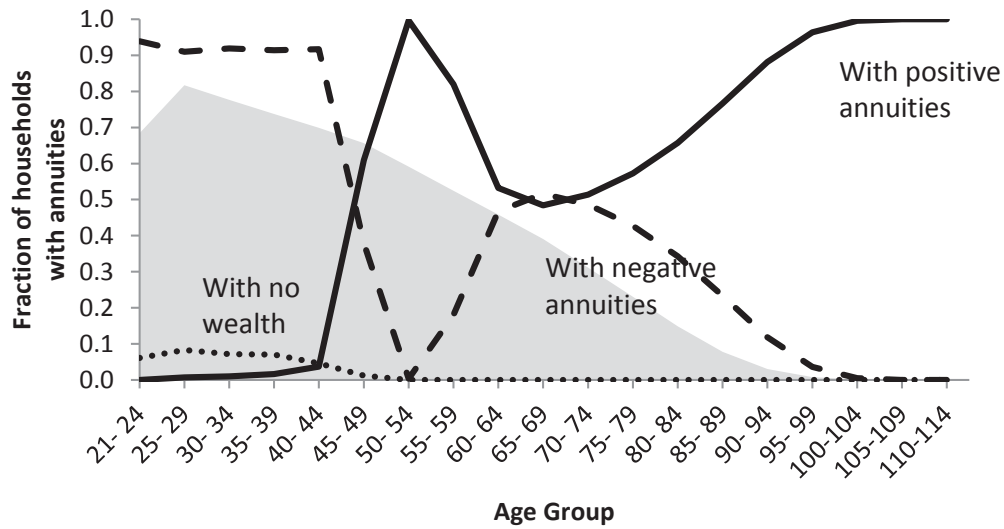
Figure 12 shows the amount of total wealth annuitized by age group, with no intentional bequest motives ($\xi = 0$) and no management fees. Notice that younger households, which tend to be healthy, hold a negative (short) position in annuities. On one hand, a negative position is costly to younger households because they must now *pay* for the mortality credit. But the mortality credit is also inexpensive when young. On the other hand, the negative annuity position provides a valuable hedge against future negative health shocks that could reduce their income before retirement and/or produce long-term care 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.²⁵ 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.

In fact, in our model the demand for negative annuities is so strong, the *aggregate* demand for annuitization across the measure of all households is actually *negative*. This result is in sharp contrast to the Yaari model, which predicts 100% positive annuitization of all wealth.

Figure 13 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

²⁵The mechanics are discussed in more detail in subsection 2.2.3. 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.”

Figure 13: 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.²⁶ 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.

Of course, a short position appears to be contrary to conventional wisdom and practice. On one hand, younger households *do* hold substantial amounts of life insurance. On the other hand, only about 17% of individuals between ages 18 and 24 hold individual life insurance policies, increasing to about 26% between ages 25 and 34 (LIMRA, 2011).²⁷ Moreover, the apparent primary motivation for buying life insurance is to protect dependents rather than to hedge future health risks (LIMRA, 2012*b*). 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.

²⁶Naturally, the shares add up to one, unless the age group has some households with no wealth. Some very poor households have no wealth and borrow against the lower bound of their future labor income produced by our model’s wage process.

²⁷These numbers increase to 36% for ages 18 to 24 and 54% for ages 25 to 34, respectively, if employer-based group life insurance policies are included.

5.3 Additional Factors That Reduce Annuity Demand

Thus far, our analysis has considered the role of uninsured expenses that substantially reduce the level of annuitization in our model with stochastic mortality probabilities but leave the level of annuitization at 100% in the Yaari model. We now consider additional factors that can reduce annuitization in both frameworks. Our intent is to see how far a rational model can be pushed to produce a low level of annuitization rather than to examine the differences between our model and the Yaari model. Still, in key places, we also report the effects of these factors on the Yaari model, for comparison.

5.3.1 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).²⁸ 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 assume a differential management fee of 1% and ignore commissions and surrender fees.

We first consider the Yaari calibration of our model and a risk aversion σ of 3.0. Along the intensive margin, 73% of all wealth remains annuitized. That consists of 100% of retiree wealth that is annuitized and 65% of non-retiree wealth. Along the extensive margin, 90% of retirees hold a positive level of annuities; about 10% of retirees have no wealth to annuitize and consume only their Social Security income. In contrast, only 33% of non-retirees hold any annuities.

However, the presence of a management fee in the Yaari model tends to reduce annuitization in a fairly knife-edge manner, as discussed earlier in Section 2. For younger households, the management fee often exceeds the value of the small mortality credit that they can earn. For them, annuitization can fall from 100% to 0%, unless they are very risk-averse and highly value consumption smoothing. In contrast, many older households earn a mortality credit that is larger than the management fee. For them, annuitization remains at 100% because the annuity return net of fees still exceeds that of bonds. Indeed, our simulation results indicate a fairly strong knife-edge effect: the fraction households that hold *any* annuities is also equal to the fraction of households that *fully* annuitized (both at 43%).

In contrast, in our baseline model with stochastic health and mortality probabilities, the level of annuitization is much smaller. Along the intensive margin, only 22% of wealth is annuitized. Of that, 41% of retiree wealth is annuitized and 14% of non-retiree wealth. Along the extensive margin, 48% of retirees hold a positive level of annuities but only 13% of non-retirees hold any annuities. Moreover, the knife-edge effect is not present: Although 19% of all households hold annuities, only 7% fully annuitize. These results are summarized in Table 5.

5.3.2 Bequest Motives

Without Management Fees. Without an intentional bequest motive ($\xi = 0$), all bequests are accidental and equal about 1.3% of GDP in our baseline model with no management fees. Bequests are, of course, zero in our Yaari calibration. 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 2.5% bequest–GDP ratio. For the Yaari calibration of our model, about 67% of wealth is annuitized (versus 100% without bequests). About 90% of households hold a positive level of annuities. In

²⁸See The Motley Fool (2013) and CNN Money (2013).

Table 5: 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%	1.3%	57%	51%	60%	37%	61%	32%	24%
0.25%	1.5%	54%	49%	56%	35%	57%	31%	21%
0.75%	1.9%	40%	44%	38%	30%	51%	26%	12%
1.00%	2.5%	22%	41%	14%	19%	48%	13%	7%
1.20%	2.6%	18%	40%	9%	14%	46%	7%	7%
<i>With Intentional Bequests</i>								
0.00%	2.5%	38%	34%	40%	35%	57%	30%	21%
1.00%	3.6%	11%	23%	6%	14%	40%	9%	6%
1.00%	4.7%	8%	15%	4%	10%	27%	7%	5%

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.

contrast, in our baseline model with stochastic mortality probabilities, only 38% of wealth is annuitized (a decrease from the 57% shown in Section 5.1.2) and only 35% of households hold a positive level of annuities.

With Management Fees. We also ran simulations that combined the same level of altruism ξ that produced a 2.5% bequest–income ratio with a 1.0% management fee. That combination increases the bequest–income ratio to 3.6%, which is consistent with the estimate by Auerbach et al. (1995). In our baseline model, the amount of wealth annuitized in the economy dropped to 11%, with only 14% 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 4.7%. To be sure, this bequest–GDP ratio might be viewed on the high side. However, this value actually matches the “lower bound” estimated by Gale and Scholz (1994) for households in the Survey of Consumer Finances. 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% (assuming a capital–output ratio of 2.8). Along with a 1.0% differential management fee, only 8% of wealth is now annuitized in our model and only 10% of households hold any annuities.

5.3.3 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 much 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.

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.

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, 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 Section 5.3.3. 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).

Appendix A: 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 can diminish 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. If agents also receive updates about their survival probabilities (and can die) at even a higher frequency, then annuities might still not dominate. Indeed, one can interpret our three periods as representing an interval of length κ 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. Full insurance not only includes protection from medical costs and lost wages, but even protection against shocks *uncorrelated* with survival, because they could still happen at the same time as a negative shock to survival probabilities, thereby producing large marginal utility. The model therefore would predict increased annuitization if such insurance markets were developed.²⁹ For *impatient* households, additional payments would also need to be made based on non-observable costs in order to directly undo changes in the annuity resale value. 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).

²⁹For example, Warshawsky, Spillman and Murtaugh (2002) consider a contract that integrates annuities with long-term care costs, but they also raise questions about regulatory hurdles and whether such a contract fully addresses the reasons we do not have a robust long-term care market.

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.³⁰ Most of the annuity literature has focused on such a contract, which is our focus as well. It is straightforward, however, to construct 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.³¹ 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.³² 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.³³

Appendix B: 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, η, 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 (η_{j+1}, h_{j+1}) and age $j + 1$. We thus use interpolation by Schumaker’s shape-preserving quadratic splines (Judd, 1998, pp. 231–234), which preserve local concavity/convexity of V as a function of wealth, thereby avoiding artificial kinks at grid points. To reduce the potential for

³⁰Because we have no inflation in our model, we could also interpret our annuity payments as being indexed.

³¹Consider, 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 $0 < \beta < 1$.

³²In 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.

³³See, for example, Sheshinski (2007), who nicely demonstrates a welfare improvement from the introduction of this unspanned contract when annuity contracts can’t 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.

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 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 10 to 20 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 C: Euler Equation Errors

Equation (11) can be rewritten more compactly as

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

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

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

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

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

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

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

or

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

Solving for consumption, we get

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

We now define the Euler Equation Error ε as

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

or

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

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 .

References

- 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.
- 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.
- 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 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. 2012. "Policy Options for the Social Security Disability Insurance Program." <http://goo.gl/22wNo>.

- 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.
- Fidelity Benefits Consulting. 2012. "Fidelity Estimates Couples Retiring In 2012 Will Need \$240,000 To Pay Medical Expenses Throughout Retirement." *Inside Fidelity*, <http://goo.gl/J40XT>.
- 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.
- 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."
- Gottlieb, Daniel. 2012. "Prospect Theory, Life Insurance, and Annuities." The Wharton School Working Paper.
- 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." *article*.
- LIMRA. 2012b. "Facts from LIMRA: Life Insurance Awareness Month, September 2012." *brochure*.
- Lockwood, Lee M. 2013. "Incidental Bequests: Bequest Motives and the Choice to Self-Insure Late-Life Risks." Northwestern University Working Paper.
- 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*.
- MetLife. 2010. "The 2010 MetLife Market Survey of Nursing Home, Assisted Living, Adult Day Services, and Home Care Costs." *Waltham, Massachusetts*.

- Modigliani, Franco. 1986. "Life Cycle, Individual Thrift, and the Wealth of Nations." *American Economic Review*, 76(3): 297–313.
- 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.
- 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.
- 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. "Population in the Social Security Area." *Actuarial Publications*, <http://goo.gl/0Sb3t>.
- 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>.
- Society of Actuaries. 2000a. "RP-2000 Female Disabled Retiree Table." <http://goo.gl/oo5F7>.
- Society of Actuaries. 2000b. "RP-2000 Male Disabled Retiree Table." <http://goo.gl/TepQI>.
- Starc, Amanda. 2012. "Insurer Pricing and Consumer Welfare: Evidence from Medigap." The Wharton School Working Paper.
- 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. 2011. "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.
- 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.