Health Dynamics and Annuitization Decisions: The Case of Social Security

Diego Ascarza-Mendoza[†] and Alex Carrasco[‡]

[†]Government School - Tecnologico de Monterrey

[‡]Massachusetts Institute of Technology

August 13, 2025

Abstract

Why do two out of three Americans claim Social Security benefits before reaching their Full Retirement Age? Why do even sufficiently rich people claim early very often? This paper resolves this puzzling phenomenon by extending a standard incomplete markets life-cycle model to incorporate health dynamics and bequest motives. Relative to the existing literature, health plays a broader role, affecting not only medical expenses and mortality but also directly the marginal utility of consumption. This role of health is disciplined using microdata on consumption, assets, income, and health from the Health and Retirement Study (HRS) and the Consumption and Activities Mail Survey (CAMS). The calibrated model successfully replicates the fraction of early claimers. Counterfactual exercises show that health-dependent preferences and bequest motives are crucial for this result. The model's success is explained by a novel channel that comes from the interaction between the negative effect of worsening health on the marginal utility of consumption, the downward health trend because of aging, and bequest motives. These two elements reduce the gains from delaying by 1) making individuals more impatient and 2) increasing the strength of bequest motives relative to future consumption. The results suggest that governments aiming to insure against longevity must consider the complementary interaction between individual incentives to insure against longevity and health risks.

Keywords: Health, Marginal Utility, Frailty Index, Social Security, Annuities

1 Introduction

The Social Security system is the primary source of government budget outlays in the United States. Due to lower fertility rates and an aging population, its sustainability is a major concern, making it a top priority for reform. One of the main goals of Social Security is to provide insurance against longevity risk to the elderly. However, individuals seem hesitant to insure against longevity, as they show low levels of annuitization. The need for Social Security reform and low levels of annuitization highlight the priority of understanding incentives to insure against longevity risk among individuals, which is the goal of this paper.

In this paper, we argue that health dynamics play a crucial role in annuitization decisions when declining health diminishes the joy derived from consumption. Our central claim is that as individuals age and their health deteriorates, the utility they derive from consumption decreases, thereby weakening the incentives to purchase annuities. This effect operates through two primary mechanisms. First, the value of annuities is closely linked to how individuals discount the future. Anticipating a decline in health, individuals expect reduced enjoyment from future consumption, which leads to increased impatience and a lower valuation of future income streams. As the continuation value of life decreases, the appeal of annuities—which provide income in later life—diminishes. Second, when individuals have concave utility over bequest motives, as is commonly assumed in the literature, they face an additional trade-off. Mortality risk creates an incentive to retain assets in each period in order to leave a bequest in the event of death. Purchasing an annuity, however, involves exchanging current assets for future income, potentially reducing the resources available for bequests. Because concave bequest preferences imply a desire to maintain asset holdings in every period, annuitization imposes a cost by generating a more uneven asset path over time. This cost becomes even more pronounced when individuals expect to be in poorer health in the future and place less value on consumption in old age. In such cases, the reduction in future consumption further discourages annuitization, as it leads to even greater fluctuations in asset holdings.

Individuals can demand annuities from the private market and government-provided annuities from Social Security. We focus on the latter. The Social Security rules establish that individuals have the option to claim benefits at different ages, between 62 and 70. The longer they wait to claim benefits, the higher the benefits they receive. In other words, when individuals delay their claiming for one year, they forgo one year of benefits in exchange for a stream of extra income while still alive. Studying Social Security claiming decisions helps us to understand the incentives for the demand for annuities. One of the advantages of focusing on Social Security is that this is an almost universal program, exempt from market frictions such as adverse selection, making it suitable to test the quantitative relevance of our channels.

Whether an annuity is attractive or not depends partly on the rate at which individuals discount the future and their life expectancy. Pashchenko and Porapakkarm (2022) finds that the Social Security annuity is quite generous up to 64 years, and then it becomes actuarially unfair assuming a discount rate of 2%. Through the lens of a standard life-cycle model that features mortality risk and this discount rate, individuals who live long enough and can afford to delay the collection of benefits should wait at least until the Full Retirement Age (FRA)[1] to claim. In the data, this constitutes the majority of people (Altig et al. (2023)). However, data from the Health and Retirement Study (HRS) shows that 45.6% of Americans born between 1930 and 1937 claimed Social Security benefits at 62 and 66% before reaching their FRA, which was 65. Interestingly, this pattern is quite independent of wealth. The early claiming behavior is consistent with a lack of demand for annuities and inconsistent with the predictions of a standard framework. The goal of this paper is to address the contrast between these model predictions and the empirical evidence.

This paper extends a standard life cycle model with mortality risk to account for the observed early claiming behavior. Our extensions consist of adding a rich characterization of the dynamics of health, the negative effect of health deterioration on the marginal utility of consumption [2], and bequest motives. Our main result is that these extensions allow us to reproduce the fraction of early claimers and most of the share of claimers at 62. The result follows from health deterioration with age, which, together with negative health dependence, reduces the valuation of additional consumption with age. This effect makes individuals more impatient and reduces their appreciation towards additional future retirement benefits. Also, because of bequest motives, as consumption loses value, preferences shift towards bequests. Since delaying implies consuming one's assets during waiting periods, the interaction between health dependence and bequest motives increases the utility cost of giving up on assets during the periods where individuals are allowed to claim.

[1] Full Retirement Age is the age at which individuals can start receiving 100% of their retirement benefits. Retirement benefits are a function of average past labor earnings. The FRA depends on the birth year of individuals

[2] Negative health-dependence is a result of our calibration exercise.

Our attention to health is motivated by two empirical findings using data from the HRS and measuring health with a frailty index as in Hosseini et al. (2022). First, unlike demographic factors such as sex, marital status, or even wealth, claiming behavior varies noticeably by health status. At every eligible claiming age, individuals in poorer health are more likely to claim Social Security benefits than those in better health, suggesting that current health conditions influence claiming decisions. Second, we find that health expectations also matter: individuals who anticipate greater health deterioration are more likely to claim early, especially if they are already in poor health. These findings motivate us to extend the standard model to incorporate a broader set of channels through which health can affect claiming behavior.

To illustrate the mechanisms, we develop a simple two-period model with non-short-selling constraints in which individuals choose how much wealth to annuitize. The framework is broad enough to encompass the Social Security claiming decision, which can be viewed as a choice to demand a public annuity. The model yields three main insights. First, when preferences depend negatively on health and health is expected to decline, individuals have less incentive to annuitize. Because annuities pay only while the individual is alive, declining health reduces the value of future consumption and makes annuities less attractive. Second, when bequest motives are present and the utility of bequests is concave, individuals face a trade-off between securing resources conditional on survival and maintaining a smoother profile of bequeathable assets. Additional annuitization reduces incidental bequests if death occurs early but increases terminal bequests if survival occurs. Third, the interaction of bequest motives with negative health-dependent preferences further discourages annuitization, as declining health lowers the desire to consume and increases the jump in assets in the survival state. In the Social Security context, the indivisibility of the annuity, the presence of short-selling constraints and market incompleteness[3] reinforce this effect: when the marginal utility of second-period consumption is sufficiently low, individuals may prefer to reduce annuitization, but the inability to do so can make early claiming the optimal choice.

With these empirical findings and a simple framework as a guide, we next turn to build a quantitative life-cycle model for the elderly with health-dependent preferences and bequest motives. Our quantitative exercise is aimed at testing whether our channels can account for the early claiming behavior of Social Security benefits, as this program is nearly universal and market

[3]With complete markets, the decision of claiming is more trivial as individuals can choose a combination of life-insurance and annuities to get a balance and make use of the good deal from delaying in the actuarial sense.

frictions-free. The model features idiosyncratic risks in income, medical expenses, and mortality. Individuals in this framework choose when to start collecting Social Security benefits and make saving-consumption decisions. We model health (frailty) as an exogenous process with rich dynamics, allowing for a deterministic and stochastic component. The deterministic component captures the effect of age on health (downward health trend). The stochastic component allows us to capture ex-ante heterogeneity, persistent shocks, and transitory shocks. With this rich characterization, health in our model may affect claiming decisions through medical expenses, mortality, and the marginal utility of consumption. The literature has explored the effect of health in claiming decisions through medical expenses and mortality, finding no significant role to account for early claiming (Pashchenko and Porapakkarm (2022)). These findings, however, contrast with the high sensitivity of claiming behavior to health in the data. Also, as shown in Blundell et al. (2020), most consumption fluctuations after retirement due to health shocks come from its negative effect on the marginal utility of consumption. Adding health-dependent preferences allows us to be consistent with the high sensitivity of claiming behavior to health changes and with the drivers of consumption fluctuations.

To discipline the model, we target different percentiles of the asset profile over the life cycle and the consumption response to transitory frailty and transitory income shocks. The asset profile is informative about bequest motives. The response of consumption to these transitory shocks is informative about the direction and magnitude of health dependence (see, for instance, Russo (2022) and Blundell et al. (2020)). The calibrated model reproduces the fraction of early claimers in the data. To evaluate what factors account for this result, we perform counterfactual experiments, adding health dependence, out-of-pocket medical expenses, and bequest motives. Our analysis shows that without these factors, the model only produces 20% of claimers at 62 (45.6% in the data) and 39% of early claimers (66% in the data). By adding bequest motives, the model produces 28% of claimers at 62 and almost 60% of early claimers. By further adding health-dependent preferences, the model produces 36% of claimers at 62 and 66% of early claimers. In line with previous results, medical expenses do not affect early claiming. The last result is because the role of medical expenses depends on their timing. Medical expenses happen very late in life, and therefore, they do not provide incentives to claim early.

The findings of this paper have significant policy implications. Firstly, it suggests that the incentives to insure against health risks and longevity are complementary. This is because people are less likely to insure against longevity if they expect a poor quality of life due to deteriorating health conditions. Therefore, governments that aim to insure their citizens against longevity risks need to complement public pensions with health coverage. Additionally, since people have strong bequest motives, it is crucial to determine whether there are potential efficiency gains for governments that aim to insure their populations against longevity risks. These efficiency gains can come from a revenue-neutral reform, which includes a shift in policy that increases life insurance provision. By providing life insurance, individuals will have more incentives to insure themselves against both longevity and health risks. Therefore, efficiency gains can result from providing insurance against only one risk, rebalancing the assigned budget to different types of insurance, or both.

The remainder of the paper is organized as follows. In section **??**, we present related literature and my contributions. Section 2 presents our empirical analysis of health and claiming decisions. Section 3 develops a simple framework to understand how health-dependent preferences affect claiming decisions. Section 4 presents our quantitative model. Section 5 presents our calibration strategy and calibration results. Section 6 shows the quantitative model's performance to account for early claiming and counterfactual experiments. Section 7 concludes.

2 Empirical Analysis

Social Security benefits are a function of average past labor earnings. This amount of benefits that correspond strictly to earnings is called the Primary Insurance Amount (PIA). Individuals are entitled to receive their PIA when they reach their FRA, a function of the year they were born. For individuals born before 1938, the FRA was 65. However, individuals can start collecting benefits at the age of 62, or they can delay the collection of benefits until they reach 70 years. The more an individual waits, the higher those benefits. While claiming at 62 provides 80% of PIA, claiming at 70 provides 132.5% of PIA for this cohort. Table 1 shows the amount of benefits as a fraction of PIA. These rules show a tradeoff between the amount of benefits and the timing at which individuals start receiving Social Security income. For individuals who can afford to delay the claiming of benefits and live long enough, this decision can imply receiving more benefits in the present discounted value sense.

This section presents three key findings: first, a significant fraction of individuals claim benefits before reaching the Full Retirement Age, and this pattern remains invariant across various characteristics, including sex, marital status, and wealth. Second, there are differences in the

Table 1 Claiming Rules for the Cohort Born Before 1937

Claiming Age	62	63	64	**65 = FRA**	66	67	68	69	70
% of Full Benefits	80%	86.7%	93.3%	100%	106.5%	113%	119.5%	126%	132.5%

likelihood of claiming at each age by health status. In particular, at each age and conditional on not claiming, individuals in worse health conditions are more likely to claim benefits than those in better health, suggesting that health may play a role in claiming decisions. Third, and which is the focus of this paper, the future evolution of health helps to account for why individuals decide to claim precisely before reaching the full retirement age.

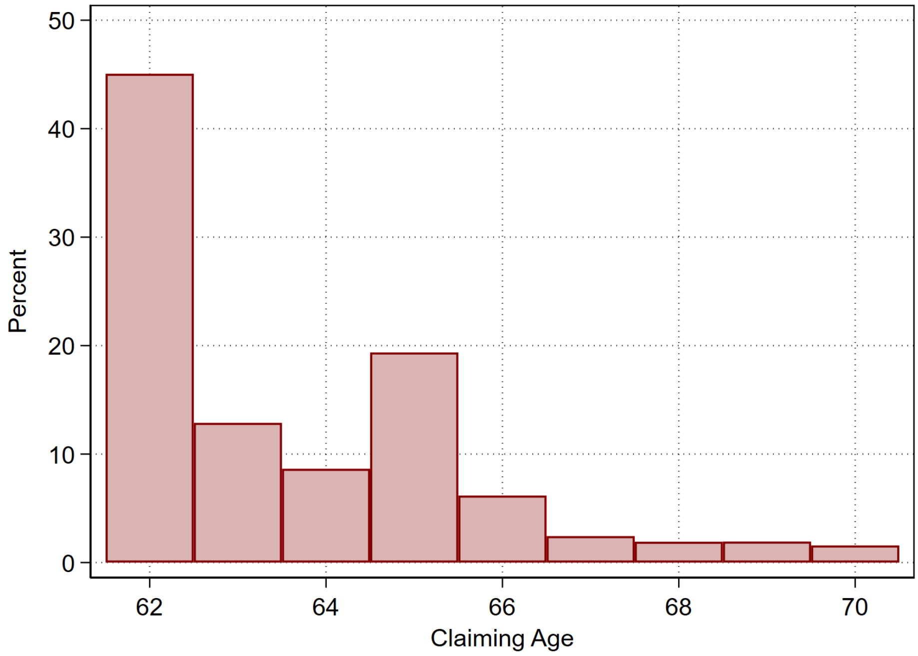
2.1 The invariance of claiming behavior by demographic characteristics

Figure 1 shows that 45.6% of individuals claim as early as possible, and at least 2 out of 3 individuals claim before the NRA. Furthermore, there is small heterogeneity across marital status, wealth, and sex, as shown in Table 2. There is more significant variation across groups by educational attainment. However, there is a significant fraction of early claimers, even with the highest educational attainment. While it is not surprising that individuals who cannot afford to delay the collection of benefits claim before reaching the Full Retirement Age do so, through the lens of a standard life-cycle model, it turns out to be puzzling that wealthy individuals, who on average live longer and can afford to delay the collection of benefits, often claim early. The puzzling nature of this phenomenon arises from the fact that with a reasonable discount rate, the Social Security annuity turns out to be generous in the actuarial sense up to the age of 64, as shown by Pashchenko and Porapakkarm (2022). From that perspective, a standard life-cycle model with a standard calibration would predict that a significantly smaller fraction of individuals should claim at 62, which is not the case. In particular, in a standard framework with mortality risk like Yaari (1965), individuals choose when to claim based on maximizing the expected present discounted value of retirement benefits. This implies that most people should wait at least until the FRA to claim.

2.2 Health and the likelihood of claiming

In this section, we document a high sensitivity of claiming behavior by health status. In particular, we show that at every possible claiming age, conditional on not claiming yet, individuals with

Figure 1 Empirical Claiming Age Distribution (cohort born before 1937)



worse health status have a higher likelihood of claiming than those with better health. To measure health, we use a frailty index following Hosseini et al. (2022). A frailty index is a number between 0 and 1 that measures the fraction of health deficits an individual has accumulated over their lifetime. In figure 2, we group individuals by age and by the quartile of frailty that they belong to[4]. The figure shows that the likelihood of claiming is quite monotonic in frailty at every age, suggesting that health can be one determinant of claiming behavior. However, this only explains why we can observe differences in claiming likelihood at a certain age due to health differences. What turns out to be puzzling, however, is why it is the case that most people choose to claim early. We believe that the latter is connected with the downward health trend that every individual faces as a consequence of aging.

2.3 Early claiming and the dynamics of health

In this paper, we hypothesize that many individuals claim early in part because health declines with age, reducing the value of annuity income: as health deteriorates, the utility derived from consumption falls. We formalize this mechanism in a simple model in the next section. To assess

[4]In this exercise, we are looking at the likelihood of claiming from 62 to 65 because most of the claiming decisions happen at those ages. Including ages after 65 has the issue of having too few observations, as most people claim at 65 or before

Table 2 Social Security Claiming Age by Demographic Characteristics

Claiming Age	62	Before 65	65
Overall	45.06%	66.58%	19.37%
Sex			
Men	44.53%	65.37%	20.15%
Women	46.05%	68.82%	17.92%
Wealth at 62			
Bottom Quintile	50.29%	73.27%	18.89%
Top Quintile	53.12%	71.83%	18.71%
Education			
Less than High School	46.59%	68.75%	18.54%
High School or GED	50.46%	71.43%	17.51%
College or Some College	39.41%	61.04%	21.50%
Marital Status			
Married or Partnered	47.16%	68.19%	19.37%
Divorced or Separated	45.06%	66.59%	19.36%
Never Married	46.29%	62.35%	22.37%

Note: The numbers are computed for individuals born in 1937 or before.

the empirical relevance of this channel, we analyze five birth cohorts observed at age 62—the earliest claiming age—and followed at least until age 70. Our proxy for expected health dynamics is the self-reported probability of experiencing health limitations that would affect work within the next ten years. Because this measure contains both noise and a systematic component related to observables (e.g., wealth, life expectancy, and recent health changes), we regress it on these characteristics and use the fitted component as our measure of expected health change.

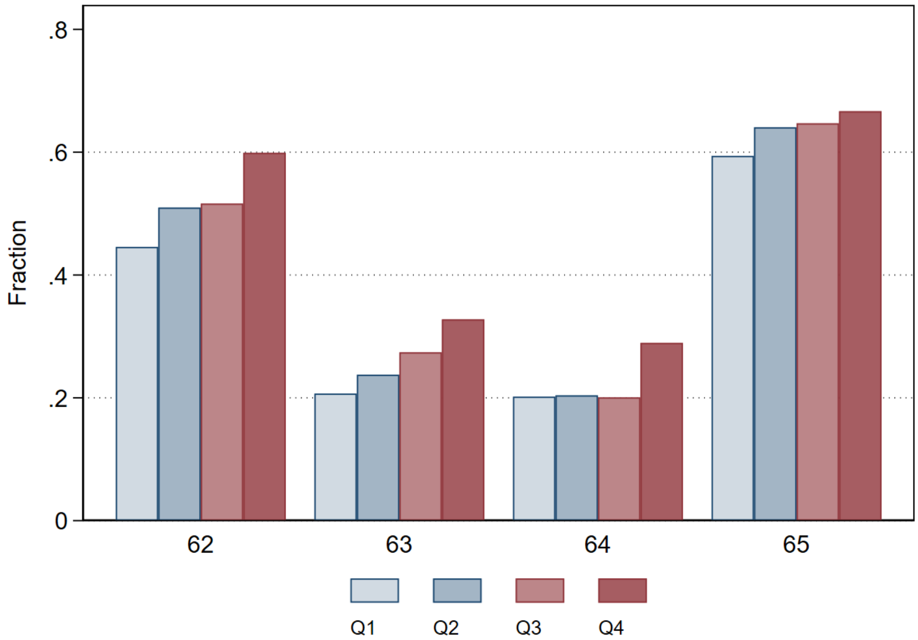
We then estimate a linear regression[5] in which the dependent variable equals 1 if an individual claims at age 62 and 0 otherwise. For ease of interpretation, we rescale this outcome to take the value 0 when the individual does not claim and 100 when they do. Our parameters of interest are the interactions between expected health change and the individual's frailty quartile at age 62. These coefficients capture how anticipated health deterioration influences claiming behavior conditional on current health status.

The specification includes controls for frailty, sex, liquid wealth (net worth excluding housing), predicted mortality risk,[6] college education, marital status, the subjective probability of leaving a

[5]Results are similar with logit and linear probability models; see the Appendix.

[6]Mortality risk is the predicted probability from a probit model with polynomials in age and frailty, sex, education, and time fixed effects.

Figure 2 : Probability of Claiming Conditional on Frailty Quartile.



Note: The figure categorizes individuals into quartiles based on their level of frailty at each specific age. For every age group, I calculate the probability of claiming by dividing the number of individuals in a particular frailty quartile who claim at that age by the total number of individuals in the same frailty quartile who have not claimed yet up to that point.

bequest, and interactions between bequest intentions and current frailty. We estimate this model separately for four measures of bequest intentions,[7] and report the results in Table 3.

The estimates are consistent with our hypothesis that anticipated health deterioration increases the likelihood of early claiming. The interaction between expected health change and frailty quartile is positive and statistically significant for most quartiles, indicating that worse expected health is associated with a higher probability of claiming at age 62. For the third quartile, however, the interaction is negative, while the frailty main effect for that quartile is positive and significant (and not significant for the others). This pattern suggests heterogeneity: among the moderately frail, expectations may matter less—either because expectation measures are noisier for this group or because current health constraints already dominate the timing decision. Bequest intentions are largely neutral at lower thresholds ("any," \$10k, \$100k), but a strong bequest motive at the \$500k threshold is associated with a lower likelihood of claiming at 62, consistent with delaying to preserve or enlarge the bequest base. Other controls behave as expected: college

[7]The HRS asks respondents for the probabilities (0–100) of leaving any bequest, more than \$10,000, more than \$100,000, and more than \$500,000.

education is associated with less early claiming, whereas coefficients on sex, liquid wealth, and predicted mortality risk are small and statistically indistinguishable from zero. Overall, the pattern—especially the positive, significant interactions in three of four frailty groups—supports the mechanism that both current health and expectations about future health meaningfully shape claiming behavior. Motivated by these findings, the next sections develop a life-cycle model that incorporates bequest motives and health dynamics and describe its quantitative implementation.

3 Theory

To explain this puzzling phenomenon, we propose a framework that incorporates four key elements: incomplete markets with borrowing constraints, concave preferences with respect to bequest motives, health-dependent utility, and anticipated health deterioration due to aging. Understanding the intuition behind this mechanism requires clarifying the role each of these components plays in shaping claiming behavior. We begin by outlining the intuition informally before presenting a formal analysis in the sections that follow.

Incomplete markets and borrowing constraints: In a setting with complete markets and no borrowing constraints, an individual's problem can be characterized by a standard intertemporal budget constraint. Within this framework, the optimal claiming decision is the one that yields the most favorable intertemporal consumption path. This typically implies that delaying benefit collection is always optimal, as it allows for higher future payouts. However, this prediction is at odds with empirical evidence, which shows that many individuals claim benefits earlier than what this model would suggest.

Bequest motives: In the presence of mortality risk, incomplete markets, and borrowing constraints, concave preferences over bequests can lead individuals to claim benefits earlier. Delaying benefit collection increases future annuity income but requires the individual to draw down assets in the present. This trade-off reduces the potential for incidental bequests during the intervening period, even if it allows for larger terminal bequests later on. In effect, each additional unit of annuity income comes at the cost of diminished incidental bequests. When preferences over bequests are concave—as is commonly assumed in quantitative analyses (see Lockwood (2018); De Nardi (2004))—individuals derive utility from smoothing asset holdings

Table 3 Claiming likelihood and health expectations

	Any	10k	100k	500k
4 quantiles of rfrailty=2	-43.88	-50.53	-63.44**	-60.97*
	(42.87)	(32.13)	(31.53)	(31.39)
4 quantiles of rfrailty=3	195.49***	199.23***	202.14***	203.75***
	(37.48)	(28.12)	(27.04)	(26.60)
4 quantiles of rfrailty=4	36.22	27.61	12.70	16.44
	(45.07)	(33.50)	(32.84)	(32.67)
4 quantiles of rfrailty=1 × Linear prediction	2.02***	2.26***	2.22***	2.44***
	(0.65)	(0.50)	(0.50)	(0.50)
4 quantiles of rfrailty=2 × Linear prediction	3.24***	3.53***	3.61***	3.62***
	(0.79)	(0.60)	(0.61)	(0.62)
4 quantiles of rfrailty=3 × Linear prediction	-2.75***	-2.56***	-2.66***	-2.65***
	(0.57)	(0.43)	(0.43)	(0.43)
4 quantiles of rfrailty=4 × Linear prediction	1.12*	1.32***	1.35***	1.37***
	(0.57)	(0.43)	(0.43)	(0.43)
Sex	3.27	-0.12	0.03	0.09
	(3.68)	(2.74)	(2.79)	(2.78)
Liquid Wealth	-0.05	0.07	0.08	0.25
	(0.37)	(0.33)	(0.32)	(0.30)
Mortality Risk	-36.14	-21.94	-23.29	-25.90
	(41.49)	(27.15)	(27.51)	(27.61)
Has college	-5.97	-8.64***	-7.10**	-5.69*
	(4.15)	(3.22)	(3.29)	(3.25)
Is married	6.48	4.47	4.92	5.32*
	(4.28)	(3.24)	(3.22)	(3.20)
Bequest (likelihood)	0.12	0.08	-0.02	-0.20***
	(0.12)	(0.08)	(0.06)	(0.07)
Bequest × Frailty (Q1)	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)
Bequest × Frailty (Q2)	-0.15	-0.10	-0.02	0.14
	(0.18)	(0.12)	(0.09)	(0.10)
Bequest × Frailty (Q3)	-0.04	-0.05	-0.10	-0.00
	(0.16)	(0.12)	(0.09)	(0.12)
Bequest × Frailty (Q4)	-0.18	-0.14	0.02	0.14
	(0.15)	(0.11)	(0.09)	(0.13)
Constant	-45.67*	-48.65**	-39.96**	-44.62**
	(26.31)	(20.03)	(19.73)	(19.49)
Observations	788	1368	1349	1339

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

over time, including for potential unplanned bequests. Consequently, the desire to maintain a smoother bequest profile can discourage delaying benefit claims. Depending on factors such as annuity returns, survival probabilities, and risk aversion, the decision to delay benefits may no longer be optimal.

Health-dependent preferences and deteriorating health: When the marginal utility of consumption declines with deteriorating health, individuals derive less utility from consumption as they age and their health worsens. This feature reduces the continuation value, making individuals effectively more impatient. As a result, they exhibit a weaker incentive to accumulate assets, including annuities and delayed Social Security income. Moreover, the presence of negatively health-dependent preferences, combined with a declining health trajectory, amplifies the role of bequest motives in discouraging annuitization. Poorer expected health in the future reduces the desirability of future consumption, leading individuals to value immediate resources more highly. Consequently, relative to a setting without health-dependent preferences, delaying benefit collection leads to larger asset holdings late in life—further impairing the ability to smooth utility from incidental and terminal bequests. This mechanism is further illustrated in the toy model presented below.

3.1 Toy Model

Assume an individual can live up to 2 periods. The probability of surviving to the next period is P. An individual wants to consume a fixed sequence of consumption. For instance, suppose an individual wishes to consume $\overline{c}$ in periods 1 and 2 (without losing generality). Suppose also that the individual has an initial wealth endowment given by w. For a matter of exposition, assume the subjective discount factor is $\beta = 1$, and the return on savings in a unique available asset is $R = 1$. The individual derives utility from consumption $u(.)$ and leaving bequests $\phi(.)$. For now, let's assume that health is absent in the setting. Then, the lifetime utility is given by:

$$v(d) = \underset{d}{Max} \quad u(\overline{c}) + \beta P[u(\overline{c}) + \beta\phi(b_2)] + \beta(1-P)\phi(b_1).$$

Where $d = 1$ denotes delay, and $d = 0$ denotes claim early. $y_t(d)$ is the Social Security income in period t given the decision of claiming d. By claiming early, the individual receives Social Security income equivalent to y units of consumption in period one and y units of consumption in period two. If the individual decides to delay, she does not receive Social Security income in the first period, and she receives an income equivalent to $(1 + R_{ss})y$ units of consumption in period 2. We will assume that w is big enough so the individual can actually afford to delay. We are interested in the case in which delaying gives more resources than claiming early:

$$y_t(d) = \begin{cases} (1-d)y & \text{if } t = 1 \\ (1 + R_{ss}d)y & \text{if } t = 2 \end{cases}$$

with $R_{ss} > \frac{1}{P}$. In this simple example, delaying implies leaving fewer bequests:

$$b_1(d=1) = w - \overline{c} < w + y - \overline{c} = b_1(d=0),$$

but implies leaving larger bequests in the second period:

$$b_2(d=1) = b_1(d=1) + (1 + R_{ss})y - \overline{c} = w + (1 + R_{ss})y - 2\overline{c}, \tag{1}$$

$$b_2(d=0) = b_1(d=0) + y - \overline{c} = w + y - 2\overline{c}. \tag{2}$$

Clearly, $b_2(d=1) > b_2(d=0)$, and importantly, by delaying the gains in additional voluntary bequests are larger than the losses in incidental bequests:

$$b_2(d=1) - b_2(d=0) = R_{ss}y > y = b_1(d=0) - b_1(d=1) \tag{3}$$

Whether delaying is optimal or not depends on the concavity of ϕ. To see why, compute the difference in the lifetime utility from delaying and the one from claiming early $\Delta_{\text{w/o health}} \equiv v(1) - v(0)$:

$$\begin{aligned} \Delta_{\text{w/o health}} = v(d=1) - v(d=0) = \underbrace{(1-P)\left(\phi(w-\overline{c}) - \phi(w+y-\overline{c})\right)}_{\text{<0 (losses from incidental bequests)}} \\ + P\underbrace{\left(\phi(w + (1+R_{ss})y - 2\overline{c}) - \phi(w+y-2\overline{c})\right)}_{\text{>0 (gains from voluntary bequests)}}. \end{aligned} \tag{4}$$

When $\Delta_{\text{w/o health}} > 0$, delaying is the optimal decision. When ϕ is linear, $\Delta_{\text{w/o health}}$ is always positive. However, f ϕ is sufficiently concave, $\Delta_{\text{w/o health}}$ can be negative. This means that with a concavity in preferences, there is a trade-off between having more resources (in this example, leaving more bequests in the second period) and having 'worse insurance,' which comes from the reduced gains because of the concavity in preferences. Next, let's add health-dependent preferences and a downward trend in health. As we will see, this addition amplifies the effect that arises from concavity. In particular, suppose now that an individual is young and in good health in the first period, and he is old and in bad health in the second period (health deterioration because of aging). Let's assume that because of a low marginal utility of consumption in bad health, the individual wants to consume $\overline{c}$ units in the first period and does not want to consume at all in the second period (for simplicity, let's assume that the utility of consuming 0 exists and it is bounded). In this case, the bequests in the first period will be as before. However, bequests in the second period will be given by:

$$b_2(d=1) = b_1(d=1) + (1+\theta)y = w + (1+R_{ss})y - \overline{c}, \tag{5}$$

$$b_2(d=0) = b_1(d=0) + y = w + y - \overline{c}. \tag{6}$$

In this case, the bequests that are left at the end of the second period are even larger than in our example without health. With concave preferences, this implies that the gains in terms of utility from having additional bequests are even smaller than before. In particular, let's define again the difference in lifetime utility:

$$\begin{aligned}\Delta_{\text{w health}} \equiv v(d=1) - v(d=0) = (1-P)\underbrace{(\phi(w-\overline{c}) - \phi(w+Y-\overline{c})}_{<0} \\ + \underbrace{P\left(\phi(w+(1+R_{ss})y-\overline{c}) - \phi(w+y-\overline{c})\right)}_{>0}.\end{aligned} \tag{7}$$

From the concavity of ϕ it follows that:

$$\phi(w+(1+R_{ss})y-\overline{c}) - \phi(w+y-\overline{c}) < \phi(w+(1+R_{ss})y-2\overline{c}) - \phi(w+y-2\overline{c}). \tag{8}$$

Which implies,

$$\Delta_{\text{w health}} < \Delta_{\text{w/o health}}. \tag{9}$$

This illustrates the main mechanism of this paper. A downward trend of health and negative health-dependent preferences reduce the demand for consumption in the second period. Because of this, the gains that come from having additional resources in the second period decrease. In other words, when the downward health trend and health-dependent preferences are added to the model, the effect of concavity is amplified. Notice, however, that if the individual had, in addition, access to life insurance or annuities so the market is complete, the optimal decision would consist of delaying.

With this intuition, we formalize the framework guiding my quantitative exercise. To understand better the role of market incompleteness, we start by describing a complete market setting in which individuals have access to bonds and annuities. In this kind of setting, the main result is that the optimal claiming age is the one that maximizes the present discounted value of benefits (delay) and that health dependence does not matter for claiming decisions. However, once markets become incomplete, even if delaying the collection of benefits might provide higher benefits than claiming early in the present-discounted value sense, frailty becomes one determinant of claiming decisions.

In particular, lousy health prospects reduce how appealing future consumption is. With a lower marginal utility of consumption in period two, the continuation value decreases, reducing the gains from more resources in the second period. Furthermore, bequests become stronger because the marginal rate of substitution between consumption and bequests shifts, enhancing the strength of bequest motives. Stronger bequest motives reduce the incentives for annuitization, as annuities only pay off as long as an individual is alive. Thus, if the effect of negative health-dependent preferences is strong enough, claiming early can be optimal because the gains from having more resources in the future become smaller than the losses from giving up on resources in the first period. Formally, we show that under certain regular assumptions, there exists a threshold of frailty such that for frailty levels above the threshold, the optimal decision consists of claiming early.

3.2 Complete Markets

Consider a retiree that lives up to two periods with an initial level of frailty f_1, exogenous initial wealth w, and a base level of Social Security benefits (y). The utility is derived from consumption ($u(c, f)$) and from leaving bequests ($\phi(b)$). Also, utility from consumption is affected by frailty (f). Frailty in the second period is given by f_2 and is fully anticipated by the agent. As a matter of exposition, we assume in this section that frailty only affects decisions through its effect on the marginal utility of consumption. The individual is alive in the second period with probability $P \in (0, 1)$. Wealth finances consumption and savings. To consider a complete market setting, we assume that the individual has access to bonds and annuities [8]. Prices of these assets are exogenous and are denoted by p^a and p^b, respectively.

The agent decides when to start collecting Social Security benefits. She can either delay the collection or start collecting them in the first period. The delay decision is denoted by $d \in \{0, 1\}$ where $d = 1$ denotes delay. Given y, the Social Security benefits are as follows:

$$y_t(d) = \begin{cases} (1-d)y & \text{if } t = 1 \\ (1 + R_{ss}d)y & \text{if } t = 2 \end{cases}$$

As can be seen, if $d = 0$, the individual receives a stream of income equal to y in both the first and the second periods. However, if $d = 1$, the individual receives 0 in the first period and $(1 + R_{ss}d)y$ in the second period. We are interested in the case $R_{ss}p^a > 1$, which means delaying provides a higher present discounted value of benefits. Given these primitives, the optimization problem is:

$$v(w, y; \boldsymbol{f_1}, \boldsymbol{f_2}) = \max_{\{c_t,\, a_t,\, b_t\}_t,\, d} \quad u(c_1, f_1) + \beta P \left[u(c_2, f_2) + \beta\phi(b_2)\right] + \beta(1-P)\phi(b_1)$$

$$s.t.$$

$$c_1 = w + y_1(d) - p^a a_1 - p^b b_1, \qquad c_2 = a_1 + b_1 + y_2(d) - p^a a_2 - p^b b_2,$$

$$c_1 \geq 0, \quad c_2 \geq 0.$$

The variable c_1 will represent consumption in the first period, while c_2 will represent consumption in the second period. We will also use the variables b_1 and b_2 to denote the quantity of bonds purchased by the individual in periods 1 and 2, respectively. Bonds can either be used to finance consumption in the second period or as bequests. Finally, we have a_1 and a_2, which represent annuity purchases made in each period. Importantly, in this setting, we will allow the possibility of issuing negative amounts of these assets. We will elaborate on why this assumption is relevant later. To keep things simple, we will be making the following set of Assumptions that we denote by **(A)**:

[8] Notice that we only need two assets to complete the market since individuals fully anticipate their health in this framework.

1. *Separability*: $u(c, f) = h(f)U(c)$ w/ $h \in C([0, f^{\max}])$, $h(0) = 1$, $\lim_{f \to f^{\max}} h(f) = 0$, and $h'(f) < 0$

2. *Regular concavity*: $U \in C$ with $U'(c) > 0$, $U''(c) < 0$, and Inada conditions

3. *Asset's pricing*: $p^b > p^a$ and $R_{ss}p^a > 1$

4. *Bequests are a luxury good*: $\phi(b) = \theta_1 U(b + \underline{b}) + \theta_2$ where θ_1, $\underline{b} > 0$ and $\theta_2 \in \mathbb{R}$

5. *Bounded luxury degree of bequests*: $\underline{b}$ is such that $u'(w + y(0), f_1) < \beta(1 - P)\phi'(\underline{b})$.

where $C([0, f^{\max}])$ denotes the space of continuous functions in the interval $[0, f^{max}]$. Without additional constraints, the individual in this problem can freely transfer resources across states and time. Thus, an intertemporal budget constraint can replace the periodic budget constraints that these individuals face. Given that the problem can be simplified by maximizing utility subject to an intertemporal budget constraint, the optimal claiming decision consists of maximizing intertemporal resources. In other words, the optimal decision is to delay. The following proposition formalizes this idea.

Proposition 1. *Under* **(A)**, $d^* = 1$.

The proof is in Appendix C.2. According to Proposition 1, in a complete market setting, frailty does not affect claiming decisions. This means that if delaying the collection of benefits results in an expansion of the lifetime budget constraint, then it is optimal to delay the collection of benefits due to market completeness.

3.2.1 Incomplete Markets

To be in an incomplete market setting, we assume the individual can only access one asset (bonds). As in the case of the complete market, we assume, for tractability, that the individual fully anticipates her frailty sequence. The optimization problem is given by:

$$v(w, y; \boldsymbol{f_1}, \boldsymbol{f_2}) = \max_{\{c_t, b_t\}_t,\, d} \quad u(c_1, f_1) + \beta P\left[u(c_2, f_2) + \beta\phi(b_2)\right] + \beta(1 - P)\phi(b_1),$$

$$s.t.$$

$$c_1 = w + y_1(d) - p^b b_1, \qquad c_2 = b_1 + y_2(d) - p^b b_2,$$

$$c_1 \geq 0, \quad c_2 \geq 0, \quad b_1 \geq 0, \quad b_2 \geq 0.$$

In this setting, the individual can not freely move resources from one state to another because of market incompleteness. Importantly, in this setting, delaying implies leaving fewer bequests in the first period in case of death and having more resources in the second period. As shown in our previous example, in this case, the optimal claiming decision is not obvious anymore. Individuals face a tradeoff between more resources in the second period and worse insurance'. This is because of the market incompleteness and the concavity of preferences. Because health dependence amplifies the effect that comes from concavity by reducing the gains from delaying, when the effect of health on the marginal utility of consumption is strong enough or an individual is fragile enough, claiming early is optimal. Before writing the formal result, let's define the following terms:

$$V(f_2, 1) = v(;,\ f_2)\Big|_{d=1}, \qquad V(f_2, 0) = v(;,\ f_2)\Big|_{d=0}, \tag{10}$$

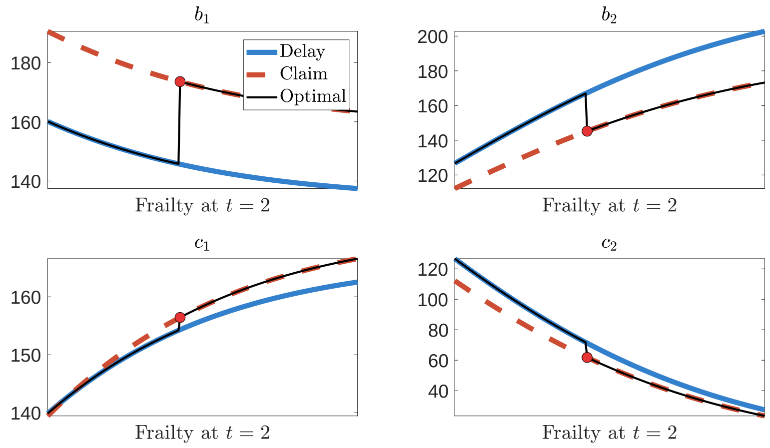
The formal result is as follows:

Proposition 2. *Suppose (A) holds, and* $\Delta \equiv V(0,1) - V(0,0) > 0$ *and finite. Then, if* Δ *is small enough or* $\left|\frac{\partial^2 u(c,f_2)}{\partial f_2 dc}\right|$ *is large enough,* $\exists f_2^*$ *such that* $V(f_2^*, 1) = V(f_2^*, 0)$ *and, for any* $f_2 < f_2^*$*, delaying the collection of benefits is optimal, while for* $f_2 \geq f_2^*$*, claiming in the first period is optimal.*

The proof can be found in Appendix C.2.

In Figure 3, we show the quantities as a function of frailty that are derived from the incomplete markets framework. The red line depicts the optimal quantities as a function of frailty in period two for an individual who decides to claim benefits in the first period. Conversely, the blue lines denote the optimal quantities conditional on delay. As discussed before, regardless of the claiming decision, with negative health-dependent preferences, the demand for consumption in the second period decreases. The figure also shows that terminal bequests also increase because preferences shift toward bequests with a lower joy of consumption. Also, because the continuation value overall decreases, individuals save less the more fragile they are and, therefore, consume more in the first period. As terminal bequests become larger, the more fragile the individual is, the gains in terms of utility from leaving larger bequests are reduced because of concavity. Furthermore, the more fragile the individual is, the more impatient she becomes, and

Figure 3 : Frailty and the optimal claiming decision.



Note: Optimal quantities as a function of f_2. The red dot denotes the threshold of frailty at which claiming in the first period becomes optimal.

therefore, the utility losses from leaving fewer bequests in case of death at the end of the first period become higher. Also, by delaying, the individual consumes less in the first period. As the gains of delaying keep decreasing, and the losses keep increasing, the individual reaches a point where she finds it optimal to claim early. In Appendix A, we show that this analysis can also be applied to study the lack of demand for private annuities.

3.3 Complete markets with borrowing constraints

In the complete market setting without short-position constraints, we discussed how the optimal claiming decision maximizes the present discounted value of benefits. We showed this for a setting that featured bonds and annuities. The market could also be completed with life insurance and bonds or with life insurance and annuities. In any scenario, the optimal rule is the same. Adding borrowing constraints, however, can affect the optimal rule in the presence of health-dependent preferences. The reason is that when the effect of health on the marginal utility of consumption is negative and big enough, the Social Security annuity may provide more consumption than wished if the individual survives. In such cases, the only way for an individual to achieve the desired level of consumption is through negative annuities or borrowing. When a borrowing constraint or a no-short position constraint becomes binding, the individual faces a tradeoff between more resources and 'worse insurance' as in the incomplete markets case. We discuss this analysis in more detail in Appendix B.

4 Quantitative Model

So far, we have developed a theoretical framework showing how negative health-dependent preferences and bequest motives can reduce the incentive to annuitize in an incomplete-markets setting. The next step is to assess the quantitative importance of these channels using a life-cycle model. Our analysis focuses on individuals' claiming decisions for Social Security retirement benefits. We choose Social Security as our example because it is a nearly universal program and is not affected by market frictions like adverse selection, which are common in the private annuity market. The following sections elaborate on the ingredients of the model.

4.1 Environment

A continuum of individuals—normalized to measure one—populates the economy. Each individual enters the model with an initial level of financial wealth, a_1, an initial frailty level, f_1, and a Primary Insurance Amount (PIA), which determines Social Security benefits if claimed at the Full Retirement Age (FRA).[9] Individuals make two key decisions: how much to consume/save and when to claim Social Security benefits. These decisions are made in the context of income, health, and mortality risks. Health is represented by a frailty index ranging from 0 to 1, with higher values indicating worse health.

4.2 Demographics

Individuals enter the model at the age of 62 and can live until the age of 99. The probability of an individual surviving to the next period depends on their state of health, as well as on age, and is denoted as $p_{t+1,t}(f_t)$. This probability is conditional on the individual being alive at period t. The frailty and age of the individual determine the probability of survival, which is consistent with the data.

[9] The PIA is a function of average past labor earnings and institutional rules. Since the focus of the model is on claiming behavior, the PIA is treated as exogenous.

4.3 Preferences

The utility in this model depends on consumption and bequests. The main feature of the preferences is that frailty affects the marginal utility of consumption. The variable c represents consumption, and f represents frailty. The parameter δ determines the direction of the health dependence in these preferences. When δ is negative, worse health conditions negatively affect the marginal utility from consumption (negative health dependence). When δ is positive, there is positive health dependence. Our estimates will show that δ is negative.

$$U(c; f) = (1 + \delta f)\frac{c^{1-\sigma}}{1-\sigma}.$$

We assume warm glow bequest motives as in De Nardi (2004):

$$\phi(b) = \phi_1 \frac{(b + \phi_2)^{1-\sigma}}{1-\sigma}.$$

The variable b denotes bequests, and σ denotes a risk aversion parameter. ϕ_1 is the parameter of the strength of bequest motives, and ϕ_2 reflects the extent to which bequests are a luxury good.

4.4 Frailty index

As mentioned before, we measure health with a frailty index. The frailty index summarizes an individual's health status, calculated by combining various deficits into one index. It is a continuous variable that allows for a straightforward interpretation of how health deteriorates with age. Additionally, the frailty index is a strong predictor of mortality, medical expenses, and nursing home admission. The frailty index represents the cumulative total of all negative health events that an individual has experienced. Generally, the frailty index combines deficits from the following categories:

- Restrictions or difficulty in activities of daily living (ADL) and instrumental ADL (IADL), such as difficulty eating, dressing, or managing money. I refer to these as ADL/IADL variables.

- Medical diagnosis or measurement such as has, or had, high blood pressure, diabetes, heart disease, cancer, or high BMI and is a current or former smoker.

- Mental or cognitive impairment such as loss of memory or mental ability or diagnosis of psychological problems. I refer to these as mental health variables.

The full set of deficits we consider to measure frailty can be found in Appendix E.1. Following Hosseini et al. (2022)[10] , we model frailty as the sum of a deterministic and a stochastic component:

$$\ln(f_{it}) = \kappa_f(t) + R_{f,it}. \tag{11}$$

Where f_{it} denotes the frailty of an individual i of age t. κ_f denotes a deterministic function of covariates that includes a polynomial in age. $R_{f,it}$ denotes the stochastic component of frailty, and it can be written as:

$$R_{f,it} = \alpha_{i,f} + z_{f,it} + u_{f,it}. \tag{12}$$

The first component, denoted by α_f, represents the individual-specific ex-ante heterogeneity in initial frailty levels. We assume that α_f follows a normal distribution across individuals with mean zero and variance $\sigma^2_{\alpha,f}$. This could be interpreted as being attributed to genetic factors. The second component characterizes the random nature of individuals' health events over their life cycles. Specifically, it comprises two elements - an autoregressive process of order one (AR(1)) and a white noise shock u_f.

$$z_{f,it} = \rho_{z,f} z_{f,it-1} + \epsilon_{zfit}. \tag{13}$$

The shocks $\epsilon_{z,f}$, and u_f are assumed to be independent of each other and independent of α_f. $\epsilon_{z,f}$ is normally distributed with mean 0 and variance $\sigma^2_{\epsilon,f}$. Furthermore, we assume that u_f is normally distributed with mean 0 and variance $\sigma^2_{u,f}$. Persistent shocks are those that affect health for more than one period—for example, an accident that causes injuries requiring years to fully heal. In contrast, transitory shocks are short-lived, such as catching the flu.

4.5 Income

In this model, individuals receive income from sources other than Social Security retirement benefits. This includes pensions, labor income, government transfers, and other sources. We denote this variable by y_{it}. We assume income to be exogenous and to have a deterministic and

[10]One difference in this framework will be that we omit the dynamics of individuals with zero frailty and instead we focus on the non-zero frailty dynamics since life in my model starts at 62 years old.

a stochastic component. The deterministic component depends on age and frailty (κ_y). The stochastic component exhibits ex-ante heterogeneity, persistent shocks, and transitory shocks, and is denoted by $R_{y,it}$. Formally:

$$\ln(y_{it}) = \kappa_y(t, f_{it}) + R_{y,it}, \tag{14}$$

$$R_{y,it} = \alpha_{i,y} + z_{y,it} + u_{y,it}. \tag{15}$$

The first component, α_y, is individual-specific and captures ex-ante heterogeneity in individuals' initial income levels. I assume that α_y is normally distributed across individuals with mean zero and variance $\sigma^2_{\alpha,y}$. This embeds, for instance, ability. The second component represents the possibility of income shocks throughout the life cycle. It is the sum of an AR(1) process and a white noise u_y.

$$z_{y,it} = \rho_{z,y} z_{y,it-1} + \epsilon_{zyit}. \tag{16}$$

The shocks $\epsilon_{z,y}$, and u_y are assumed to be independent of each other and independent of α_y. $\epsilon_{z,f}$ is normally distributed with mean 0 and variance $\sigma^2_{\epsilon,y}$. Furthermore, we assume that u_y is normally distributed with mean 0 and variance $\sigma^2_{u,y}$.

4.6 Out-of-pocket Medical Expenses

Individuals incur out-of-pocket medical expenses, denoted by $m_{i,}$, which represents the amount paid by individual i in period t. In the model, medical expenses have both a deterministic and a stochastic component. The deterministic component, $\kappa_m(t, f_t)$, depends on the individual's age and frailty level. The stochastic component captures transitory shocks and is represented by u_m , which is assumed to follow a normal distribution with mean zero and variance $\sigma^2_{u,m}$. Importantly, the randomness in medical expenses arises not only from these transitory shocks but also indirectly through shocks to frailty. The log of out-of-pocket medical expenses can be expressed as:

$$ln(m_{it}) = \kappa_m(t, f_{it}) + u_{m,it}. \tag{17}$$

4.7 Markets and governments

Markets are incomplete as individuals only have access to a risk-free asset that yields a gross return of R, which they can purchase every period. However, individuals are not allowed to borrow. It is important to note that we will not consider private annuity markets in this quantitative model for two reasons. Firstly, we want to focus on the explanatory power of the model to understand claiming decisions. This approach has the advantage of being 'frictions-free,' meaning that we do not have to deal with the complexities of other markets that affect private annuity markets, such as adverse selection and fees that create a gap between market prices and actuarially fair prices. Secondly, less than 4% of individuals hold annuities in their portfolios, making it costless to exclude annuities from our quantitative exercise.

The government in this model has the following roles. First, it pays Social Security benefits, according to each individual's exogenously given full benefits and their claiming age. In particular, it establishes a Social Security Payments function that depends on individuals' PIA and the claiming age. This can be written as:

$$SS_{it} = SS(t, PIA_i, CA_i). \tag{18}$$

The time dependence on Social Security benefits reflects the existence of minimum and maximum eligibility ages. Additionally, the government provides a consumption floor $\underline{c}$ to individuals who experience significant medical expenses. The consumption floor is designed to capture means-tested government programs such as Medicaid or food stamps. Essentially, the government provides financial assistance to individuals whose total cash in hand is insufficient to meet the $\underline{c}$ threshold.

$$Tr = Max\left\{0, \underline{c} + m - (Ra + y + SS)\right\}. \tag{19}$$

4.8 Individual Decision Problems

At the beginning of each period, individuals observe their frailty level, earnings, and medical expenses. Based on this information, they decide how much to consume and save, and whether to claim Social Security benefits if they have not already done so. To economize on notation, we denoted a subset of our state variables as:

$$X \equiv (t, PIA, a, z_f, z_y, \alpha_f, \alpha_y). \tag{20}$$

The set of state variables X is composed of age (t), the full benefits PIA, the level of assets a in the current period, the permanent components for frailty and earnings z_f, z_y, and the initial heterogeneity in frailty and earnings α_f, α_y. $V(X)$ denotes the value function of an individual who has not decided to claim benefits. Also, denote by $W^e(X, D = 0)$ the value function of an individual who decides not to claim benefits in the current period, and denote by $W^e(X, D = 1)$ the value function of an individual who decides to start collecting benefits in the current period. Finally, $V_c(X, t^c)$ denotes the value function of an individual who started collecting benefits at the age of t^c. The claiming age becomes an additional state variable for this last case.

4.8.1 Individuals who have not claimed benefits yet

Individuals who have not claimed Social Security benefits yet need to make two decisions: 1) whether to start collecting benefits or not in the current period, which we denote by the decision variable $D \in \{0, 1\}$, where $D = 1$ means the decision of delaying the collecting benefits one period, and $D = 0$ denotes the decision of start collecting benefits. 2) how much to save and consume. I denote consumption by c, and savings for the next period as a'. Consistently with the institutional rules, I also assume a maximum possible claiming age that we denoted by t^c. In other words, this problem can only apply to individuals with an age t such that $1 < t < T^c$. The optimization problem is as follows:

$$V(X) = \underset{D\in\{0,1\}}{Max}\, W^e\,(X, D)\,.$$

Where

$$W^e(X, D = 1) = \underset{c,a'}{Max} \quad U(c, f) + \beta\,\{p_{t+1}(f_t)\mathbb{E}\,[V(X')] + (1 - p_{t+1}(f_t))\,\phi(b')\}\,,$$

$$W^e(X, D = 0) = \underset{c,a'}{Max} \quad U(c, f) + \beta\,\{p_{t+1}(f_t)\mathbb{E}_t\,[V^c(X', t^c)] + (1 - p_{t+1}(f_t))\,\phi(b')\}\,,$$

$$s.t.$$

$$c + a' + m \leq Ra + \mathbb{I}(D = 0) \times SS + y + Tr,$$

$$a' \geq 0,$$

$$b' = a',$$

$$Tr = Max\left\{0, \underline{c} + m - (Ra + y + \mathbb{I}(D = 0) \times SS)\right\}.$$

The expectation here is taken over the permanent components of frailty and earnings.

4.8.2 Individuals already collecting Social Security benefits

When a person already receives Social Security benefits, the age at which they start claiming those benefits, denoted as t^c, becomes a state variable. This age determines the amount of Social Security benefits they will receive based on their PIA. In this situation, individuals only need to decide how much they want to save and spend. The optimization problem is:

$$V^c(X, t^c) = \underset{c,a'}{Max} \quad U(c, f) + \beta\left\{p_{t+1}(f_t)\mathbb{E}_t\left[V^c(X', t^c)\right] + (1 - p_{t+1}(f_t))\,\phi(b')\right\},$$

$$s.t.$$

$$c + a' + m \leq Ra + SS + y + Tr,$$

$$a' \geq 0,$$

$$b' = a'$$

$$Tr = Max\left\{0, \underline{c} + m - (Ra + y + \mathbb{I}(D = 0) \times SS)\right\}.$$

5 Calibration

There are two sets of parameters in this model. The first set consists of predetermined parameters that take standard values from the literature. The second set of parameters is calibrated by following the methodology of Gourinchas and Parker (2002) and De Nardi et al. (2010). Firstly, we estimate the parameters that govern the exogenous process of frailty, mortality, earnings, and out-of-pocket medical expenses. These parameters are estimated without using our model. Secondly, using the parameters obtained in the first step, we calibrate the remaining parameters using the Simulated Method of Moments (SMM). Specifically, we calibrate health dependence, the parameters associated with the bequest motives, and the consumption floor. We calibrate these parameters to match the evolution of wealth after retirement and the consumption response to transitory shocks in income and frailty. The definition of consumption response to transitory shocks follows Blundell et al. (2008), and Kaplan and Violante (2010).

5.1 Predetermined parameters

The parameter values for the subjective discount factor (β), the return for the unique asset (R), and the coefficient of relative risk aversion (σ) are taken from Gourinchas and Parker (2002). Specifically, β is set to 0.96, R is set to 1.02, and σ is set to 1.5.

5.2 Data

Our main data sources are the Health and Retirement Study (HRS) and the Consumption and Activities Mail Survey (CAMS). The HRS is a longitudinal survey that represents the population of the United States over the age of 50 and their spouses. It provides extensive information on health, income, assets, Social Security, demographics, and other variables. The CAMS is a supplementary study that collects data on household spending for a wide range of consumption categories, time use, and employment. It is administered to a subset of HRS respondents. We use CAMS to calculate consumption-related moments and their interactions with health and earnings.

5.3 First-Step calibration

This section shows the estimation procedure for survival probabilities, frailty, income process, and medical expenses. More details about these estimations can be found in Appendix H.

5.3.1 Frailty process:

The frailty process is estimated following Hosseini et al. (2022). The estimation of the frailty process is also done in two steps. First, We employ a Probit regression to estimate mortality probabilities. Using these estimates, we estimate the parameters of the frailty process using the

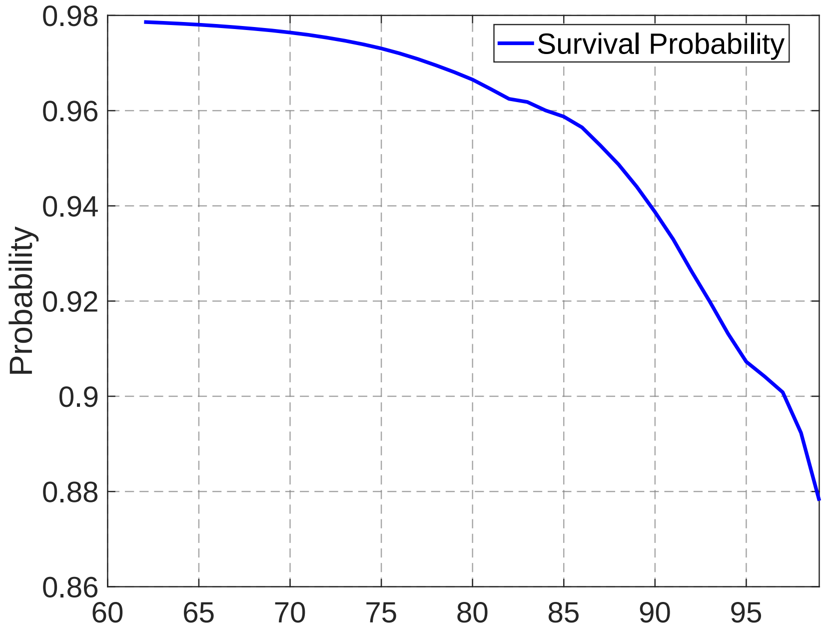
Simulated Method of Moments (SMM). The usage of SMM is motivated by the high correlation between mortality and frailty. Because there is high attrition due to mortality, which is correlated with frailty, using an auxiliary model for simulation allows me to control for potential problems due to mortality selection. The Probit regression for mortality is as follows:

$$Prob(d_{i,t+1} = 1 | d_{i,t} = 0) = \Phi(X_{it}'\gamma). \tag{21}$$

Where Φ represents the standard normal CDF. X_{it} denotes the covariates of this regression, which includes a polynomial of second order in age, a polynomial of second order in frailty, and years of education. It is worth noting that since there are no restrictions on the values frailty can take, it is possible to obtain values greater than one. In such cases, we assume that the individual dies automatically. However, it is important to note that Hosseini et al. (2022) has shown that this assumption does not affect the performance of frailty in predicting health outcomes. Figure 4 displays the calculated survival probability at the average frailty level and as a function of age. As expected, since age reduces the survival probability, and individuals also become more fragile with age, the likelihood of surviving to the next period follows a decreasing path. Importantly, our estimates and the model simulation produce a life expectancy conditional on being alive at 62 of 20 years, which is consistent with the data.

Using the estimates for mortality, we proceed to estimate the parameters governing the frailty process. To do this, we use an auxiliary simulation model specified in equation 11 and the SMM. The auxiliary model describes the dynamics of frailty. Specifically, we target the age profile of the mean of log frailty in 2-year groups between 50 and 95 years and the variance and autocovariance profile by age group. The age profile of the mean of log frailty is informative about the deterministic component of frailty that depends on age, while the variance and autocovariance profile is informative about the parameters that govern initial heterogeneity, persistence, and the volatility of the different shocks in the stochastic component of frailty. In Table 4, we show the estimates for the parameters governing the stochastic part of frailty. Consistent with Hosseini et al. (2022), we find a high persistence of frailty. The parameter governing the persistent and transitory shock is also very similar to former estimates. Our estimate of the variance of the parameter governing the fixed effect is higher than in Hosseini et al. (2022). This is because they include sex and education, which are age-invariant in their estimation. Therefore, the effect of these variables is captured in the initial heterogeneity in our setting. Figure 5 shows the

Figure 4 : Estimated survival probabilities



Note: Survival probability as a function of age evaluated at the average frailty.

estimated mean of the log of frailty compared with the data. The figure shows that the model does a good job of capturing the dynamics of frailty with slight deviations at the beginning and at the end of the life cycle. Regarding the estimates of the survival probability, the model produces an initial life expectancy equal to 20 years, consistent with the data, and a median life span of 20.7 years. In the data, the median life expectancy is 21. Overall, the model is fed with accurate estimates of frailty and survival probabilities.

5.3.2 Income process:

We measure income as all sources of income, excluding retirement benefits from Social Security. This measure allows me to maintain consistency with the model. Appendix E.4 shows all the components of this variable. To estimate the income process specified in equation (14), the deterministic function κ_y is estimated by regressing the logarithm of income on a second-order polynomial of age, frailty, education, and cohort dummies. We extract the educational component of income by subtracting the predictions from education since education is not present in the model. This clean component is then regressed on frailty and age. To estimate the stochastic

Table 4 Estimation results for the stochastic component of the frailty process

Parameter	Value
$\rho_{z,f}$	0.944
$\sigma^2_{\epsilon,f}$	0.035
$\sigma^2_{u,f}$	0.006
$\sigma^2_{\alpha,f}$	0.737

component of earnings, we compute the residuals of the previous regression and the matrix of variance and autocovariance. Using equally weighted minimum distance, we estimate the autoregressive coefficient $\rho_{z,y}$, the variance of the shock to the persistent component $\sigma^2_{\epsilon,y}$, the variance of the transitory shock $\sigma^2_{u,y}$, and the variance of the initial persistent effect $\sigma^2_{\alpha,y}$. The estimates for the stochastic component of income are shown in Table 5. Consistent with former estimates, income also exhibits high persistence. The parameters governing the volatility of the transitory shock and the fixed effect are slightly lower than previous estimates. This is because our estimation is for the elderly, who have a lower participation in the labor market.

Table 5 Estimation results for the stochastic component of income process

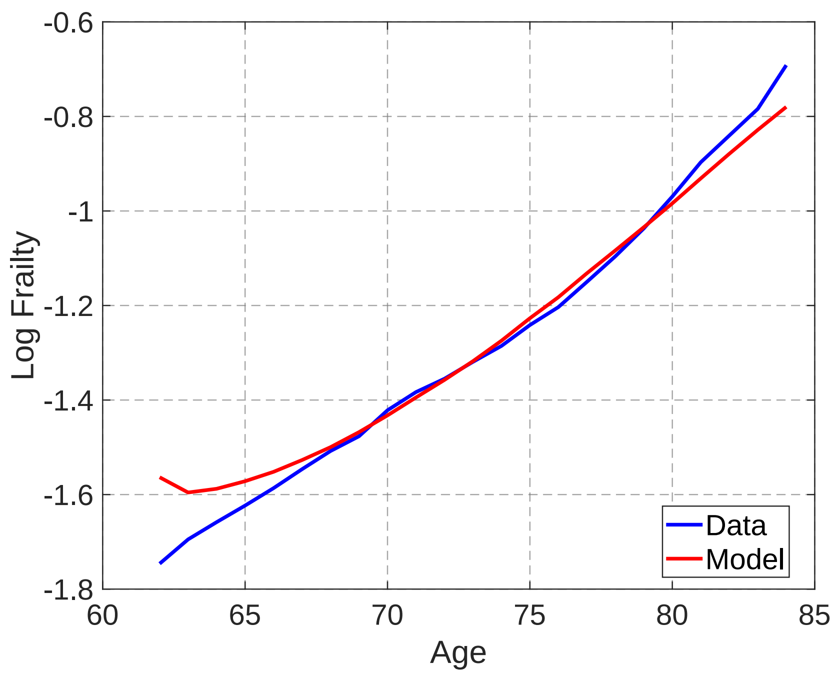
Parameter	Value
$\rho_{z,y}$	0.932
$\sigma^2_{\epsilon,y}$	0.233
$\sigma^2_{u,y}$	0.001
$\sigma^2_{\alpha,y}$	0.023

Figure 6 shows the estimates of the average log of income compared to its data counterpart. The figure shows that income follows a downward path after the fifties, consistent with reduced income because of retirement.

5.3.3 Out-of-pocket medical expenses:

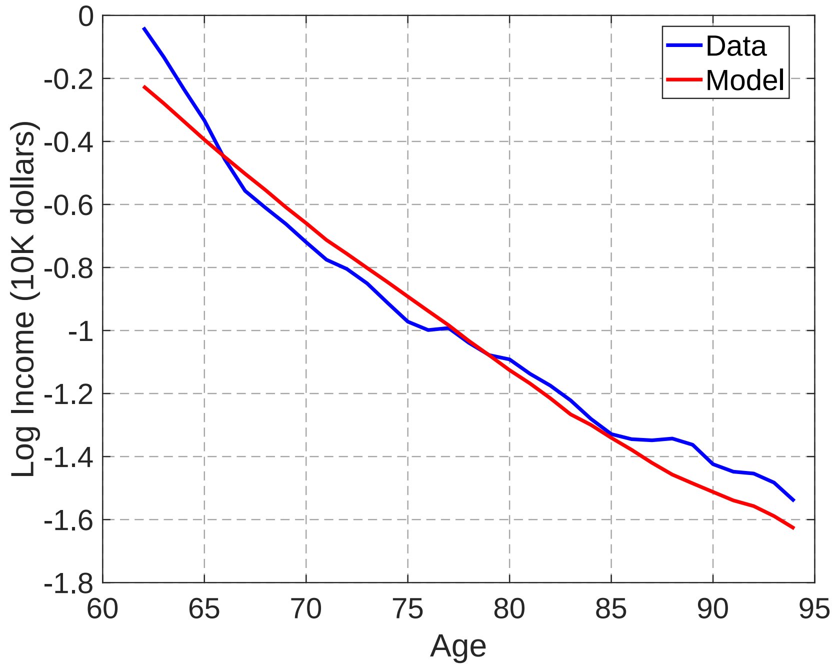
Medical expenses are the sum of what the household spends out-of-pocket on doctor visits, hospital and nursing home stays, prescription drugs, and insurance premiums. We adjust this variable by family size. Then, the deterministic function κ_m specified in equation 17 is estimated

Figure 5 : Average frailty over the life cycle



Note: Comparison of the estimated average log frailty by age with its empirical counterpart.

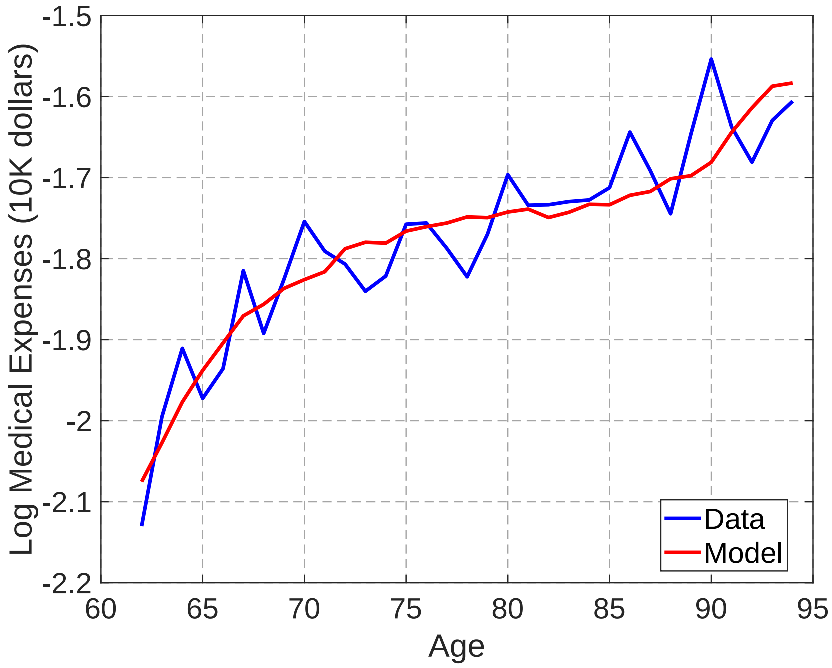
Figure 6 : Average income over the life cycle



Note: Comparison of the estimated average log income by age with its empirical counterpart.

by regressing the log of medical expenses on a second-order polynomial in household age, frailty, education level, and cohort effects. As with earnings, we extract the educational component and then regress it on a polynomial of age and frailty. To estimate the variance of the transitory shock to medical expenses $\sigma^2_{u,m}$, we regress the squared residuals from the previous regression on the same covariates. The variance of the predicted values of this last regression will then give the shock variance, which we find to be 0.073. Figure 7 shows the average of the log of out-of-pocket medical expenses compared with its estimated value. The figure shows that the estimates capture well the rapid growth of medical expenses with age. Capturing this trend well is important because the direction and the magnitude of the effect of out-of-pocket medical expenses on claiming decisions depend heavily on the timing of the expenses. In particular, when expenses happen late in life, they generate incentives to insure against longevity (delay).

Figure 7 : Average out-of-pocket medical expenses over the life cycle



Note: Comparison of the estimated average log medical expenses by age with its empirical counterpart.

5.4 Second-Step calibration

In this step, we use the parameters obtained from the first-step estimation and our model to calibrate the remaining parameters using SMM. Four parameters need to be estimated in this stage: the parameter that governs the strength of bequest motives (ϕ_1), the parameter that governs the degree to which bequests are a luxury good (ϕ_2), the consumption floor ($\underline{c}$), and the parameter that governs health dependence in preferences (δ). Specifically, we define θ as ($\phi_1, \phi_2, \underline{c}, \delta$). The estimated parameters are denoted as $\hat{\theta}$ and they satisfy:

$$\hat{\theta} = \underset{\{\theta\}}{Arg\ Min} \left\{ \left[\hat{m} - m(\theta) \right]' W \left[\hat{m} - m(\theta) \right] \right\}. \tag{22}$$

Where $\hat{m}$ denotes empirical targetted moments, while $m(\theta)$ are the analogous model-generated moments through simulation. W is a weighting matrix (set to be the identity matrix in this case). The moments we target are the following:

5.4.1 Pass-through coefficients (self-insurance)

This section follows Blundell et al. (2020) and Russo (2022). Their approach identifies health dependence by disciplining a model with the consumption response to transitory income and frailty shocks. The intuition comes from noticing that a negative income shock is equivalent to a shock of out-of-pocket medical expenses since both affect the budget constraint similarly. Because a transitory frailty shock affects not only medical expenses but also the marginal utility of consumption, by targeting the response of consumption to these two shocks, it is possible to identify the parameter governing health dependence. We apply this approach to identify health dependence in our model as the same intuition applies to our setting. The procedure is as follows: first, following Blundell et al. (2008), the self-insurance of consumption to a transitory shock x_t (pass-through coefficient) is defined as the ratio of the covariance between log-consumption growth and the shock and the variance of the shock:

$$\phi^x = \frac{cov(\Delta \ln c_t, x_t)}{var(x_t)}. \tag{23}$$

The pass-through coefficients measure the share of the variance of a shock that translates into consumption growth. In equation (23), c_t denotes non-medical consumption detrended by observables, and Δ denotes the first-difference operator. A coefficient of zero means full insurance, while a coefficient of one indicates no insurance at all. The estimation of pass-through coefficients follows a semi-structural approach in this literature. To estimate them, we first use the estimates of the parameters that govern the deterministic component of income and frailty, denoted by κ_y and κ_f, respectively. Next, we detrend income and frailty from observables and denote the detrended variables as $\tilde{y}_t \equiv \ln(y_t) - \kappa_y(t, f_t)$ and $\tilde{f}_t \equiv \ln(f_t) - \kappa_f(t)$. Also, define the quasi-differences for these variables as follows: $\tilde{\Delta} y_{t+1} = \tilde{y}_{t+1} - \rho_{z,y}\tilde{y}_t$, and $\tilde{\Delta} f_{t+1} = \tilde{f}_{t+1} - \rho_{z,f}\tilde{f}_t$. The pass-through coefficients of transitory income shocks and frailty shocks, which we denote by ϕ^y and ϕ^f, respectively, are identified using the following equations:

$$cov(\Delta \ln c_t, \tilde{\Delta} y_{t+1}) = -\rho_{z,y} cov(\Delta \ln c_t, u_t^y), \tag{24}$$

$$cov(\tilde{\Delta} y_t, \tilde{\Delta} y_{t+1}) = -\rho_{z,y} var(u_t^y), \tag{25}$$

$$cov(\Delta \ln c_t, \tilde{\Delta} f_{t+1}) = -\rho_{z,f} cov(\Delta \ln c_t, u_t^f), \tag{26}$$

$$cov(\tilde{\Delta} f_t, \tilde{\Delta} f_{t+1}) = -\rho_{z,f} var(u_t^f), \tag{27}$$

By dividing (24) with (25) and (26) with (27), we obtain:

$$\phi^y = \frac{cov(\Delta \ln c_t, \tilde{\Delta} y_{t+1})}{cov(\tilde{\Delta} \ln y_t, \tilde{\Delta} y_{t+1})}, \quad \text{and} \quad \phi^f = \frac{cov(\Delta \ln c_t, \tilde{\Delta} f_{t+1})}{cov(\tilde{\Delta} \ln f_t, \tilde{\Delta} f_{t+1})}. \tag{28}$$

We use the equations in (28) to estimate the pass-through coefficients. The equations above deliver the following moment conditions:

$$\mathbb{E}\left[\Delta \ln c_{it} \left(\tilde{\Delta} y_{it+1}\right) - \phi^y \tilde{\Delta} y_{it} \left(\tilde{\Delta} y_{it+1}\right)\right] = 0. \tag{29}$$

$$\mathbb{E}\left[\Delta \ln c_{it} \left(\tilde{\Delta} f_{it+1}\right) - \phi^y \tilde{\Delta} f_{it} \left(\tilde{\Delta} f_{it+1}\right)\right] = 0. \tag{30}$$

Given these moment conditions and following the literature, we use the Generalized Method of Moments (GMM) to estimate the pass-through coefficients of interest. Our estimates are illustrated in Table 6.

The coefficients of this study go in the same direction as Blundell et al. (2020) and Russo (2022). Specifically, the first paper indicates that the pass-through coefficient from transitory health shocks to consumption is 0.17, similar to our estimate. Our estimates suggest that a transitory increase of 0.1 in frailty results in a 1.6% decrease in consumption, and this is statistically significant with a confidence level of 99%.

Table 6 Pass-through coefficients of transitory frailty and income shocks to consumption

Parameter	Value
ϕ_u^f	-0.169^{***}
	(.062)
ϕ_u^y	0.064^{**}
	(.032)

*** p<0.01, ** p<0.05, * p<0.1

5.4.2 Asset Moments

We use twenty-four moments associated with asset accumulation. Specifically, we target the 25th, median, and 75th percentile of the wealth distribution for people in 3-year age groups between the ages 62 and 85. Assets is defined as the net worth from the HRS adjusted by family size, using the OECD household equivalence scale. To control for cohort effects, we regress this normalized variable on a set of age and cohort dummies as in Pashchenko and Porapakkarm (2022):

$$a_{it} = \sum_k \beta_k \mathbb{I}[k = age_{it}] + \sum_c \beta_c \mathbb{I}[c = \text{birth year}] + \epsilon_{it}. \tag{31}$$

Using the estimates, we compute the net worth for our cohort of interest (1937):

$$\hat{a}_{it} = \sum_k \hat{\beta}_k \mathbb{I}[k = age_{it}] + \sum_c \hat{\beta}_c \mathbb{I}[c = 1937] + \hat{\epsilon}_{it}. \tag{32}$$

This procedure allows us to include more observations from the data. Given these empirical moments, we solve the model for a given set of parameters to calculate the model-generated moments and measure the distance between those according to equation 22. Our parameters are calibrated to minimize the distance between the empirical and model-generated moments. We show technical details in Appendix H.

5.5 Second-step estimation results

In Table 7, we present the internally calibrated parameters. The table showcases that the health dependence parameter is negative. This suggests that the marginal utility of consumption decreases with worsening health conditions. The magnitude of the parameter indicates that a decline in health equal to one standard deviation in frailty decreases the marginal utility of consumption by 7.87%, provided that both consumption and frailty are at their average values. The marginal utility of consumption's reduction with higher frailty values is crucial to the claiming decision. In other words, this negative value adjusts preferences towards bequests and implies a higher discount factor than what β and the survival probabilities solely provide. Our estimates also demonstrate that bequest motives are a luxury good consistent with the savings literature. Bequest motives in this model arise from matching the asset profile of individuals. Higher bequest motives lead individuals to keep more assets. Relative to former estimates, one way to compare the strength of bequest motives is through the bequest threshold, which we denote by $\overline{b}$ and the Marginal Propensity to Bequeath (MPB). Following Pashchenko and Porapakkarm (2022), $\overline{b}$ is defined as the level of assets at which bequest motives become operative in a two-period consumption-savings decision model, and MPB is defined as the fraction of assets above the bequest threshold that individuals leave as bequests. Given the functional form for bequest motives:

$$\overline{b} = \left(\frac{\beta \phi_1}{1 - \delta f} \right)^{-\frac{1}{\sigma}} \phi_2. \tag{33}$$

$$MPB = \frac{1}{1 + \left(\frac{\phi_1 \beta}{1 - \delta f} \right)^{-\frac{1}{\sigma}}}. \tag{34}$$

In the Appendix F, we show the details of how to obtain these expressions. As can be seen, both the bequest threshold and the MPB are affected by frailty and the degree of health dependence (δ). In particular, when an individual is more fragile and the joy of consumption declines, the threshold of assets at which individuals start leaving bequests decreases, and the MPB increases. In other words, with health-dependence, the bequest motives are strengthened. To have a notion of comparability with other studies, we compute $\overline{b}$ and MPB at the average value of frailty, which is 0.21. With this value of frailty, we obtain an MPB of 0.94 and a bequest threshold of \$11,158 expressed in dollars of 2021. Former estimates from De Nardi et al. (2010), Lockwood (2018), and Pashchenko and Porapakkarm (2022) suggest that the MPB is in the range of 0.78 and 0.96, while the bequest threshold is between \$4,900 and \$21,900. Thus, our estimates are in the range of former estimates in the literature at average levels of frailty. Finally, the consumption floor in our model equals \$5,320 in 2021 dollars. This value is consistent with previous estimates in the literature, which range between \$3,000 and \$7,000.

Table 7 Preference parameters and consumption floor.

Parameter	Value
δ	-0.82
ϕ_1	50.70
ϕ_2	16.14
$\underline{c}$	'$ 5,320'

5.6 Model Fit

As indicated before, we calibrate the model to match the pass-through coefficients of frailty and earnings shocks and the asset accumulation profile of individuals. Figure 8 shows the model fit of the asset-profile accumulation implied by the model. As can be seen, the model does a good job of reproducing the asset profile accumulation. The model is more imprecise in capturing the asset profile of the top quartile, which is something the literature has also found difficult to capture.

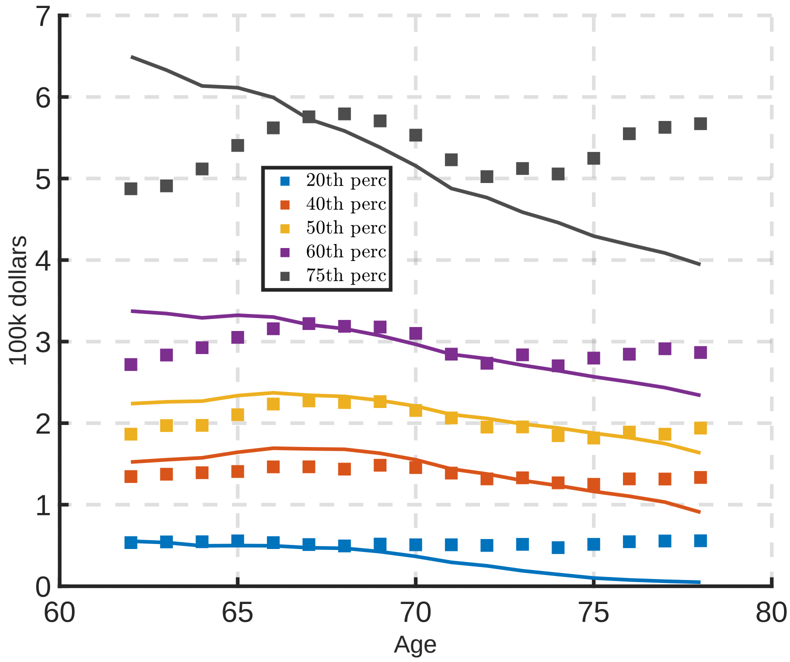
Table 8 Model fit of pass-through coefficients.

Moment	Data	Model
ϕ^y	0.064	0.049
ϕ^f	−0.169	−0.181

6 Results

In this section, we assess the performance of our quantitative model in accounting for claiming behavior, which is untargetted in our calibration. Because of this, the results should be interpreted as how much early claiming behavior can be accounted for when the model is disciplined to match the asset profile and the consumption response to transitory frailty shocks. Figure 9 compares the claiming behavior produced by the model with the empirical claiming distribution for the cohort born before 1937. The figure shows that the model can replicate the observed early claiming behavior. While the model underestimates the fraction of claimers at 62, it successfully replicates the fraction of individuals who claim before 65 (FRA). The reason why many people

Figure 8 Asset Accumulation Profile

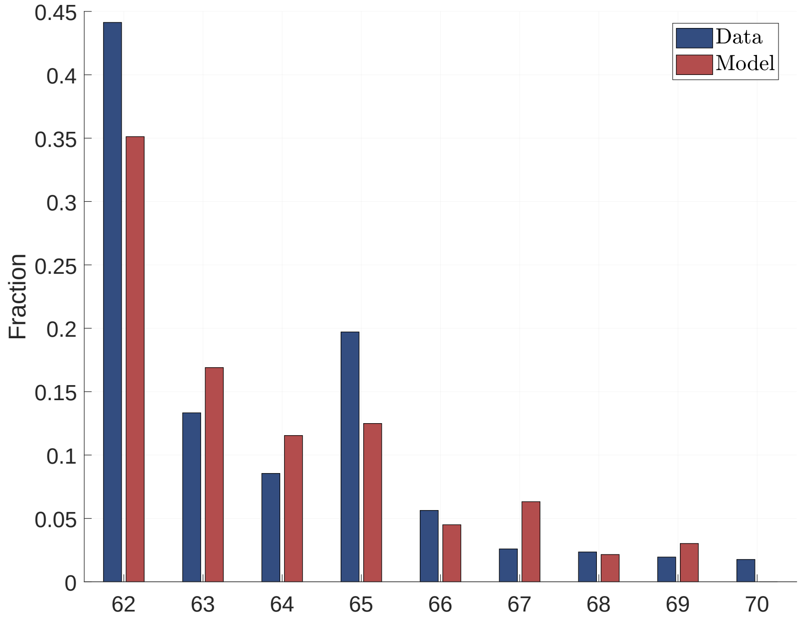


Note: Assets by 5-year age group for 20th, 40th, 60th and 75th percentiles. Dots represent data, and the lines represent the model-generated moments. The exercise contemplates the profile between 62 and 82 years.

choose to claim their Social Security benefits at age 62 is due to the way the annuity is priced and how individuals value it. According to a study by Pashchenko and Porapakkarm (2022), the Social Security annuity is not actuarially fair. In particular, while the implied pricing of this annuity is even lower than an actuarially fair up to reaching 64 years with a discount rate of 2%, this reverts afterward. With health dependence factored in, the effective discount rate becomes even higher, making the annuity seem unfair to individuals sooner.

Consistent with this idea, the following section shows that early claiming is accounted for by adding health-dependent preferences and bequest motives. In other words, a model that includes only mortality risk and out-of-pocket medical expenses cannot produce this outcome. Health-related preferences are crucial for early claiming since they raise the impatience rate and intensify the strength of bequest motives.

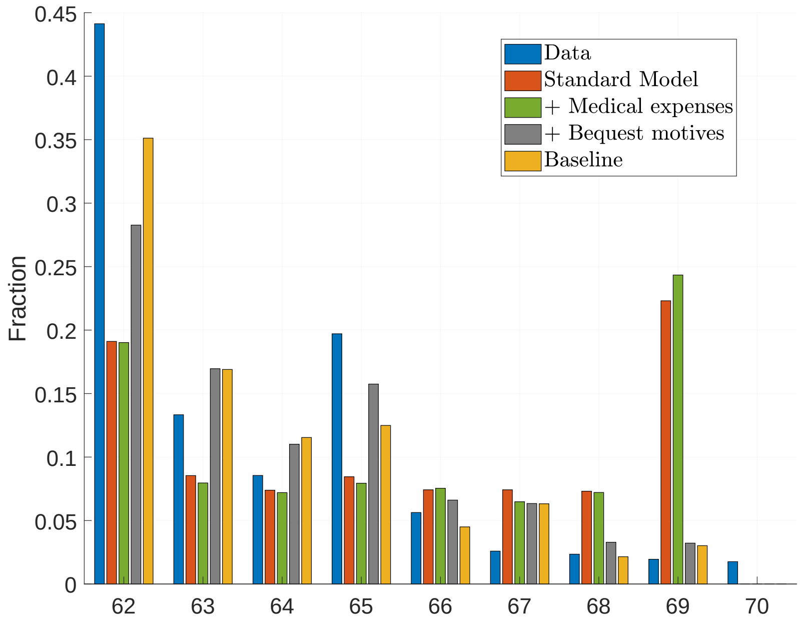
Figure 9 Baseline Model



6.1 What factors account for early claiming?

We conduct counterfactual experiments to determine what features in the model help to account for early claiming. We begin with a standard life-cycle model, which only considers mortality risk, and gradually add extensions. The results indicate that health-dependent preferences and bequest motives are the primary factors contributing to early claiming. Including out-of-pocket medical expenses has a negligible impact on claiming decisions. The findings are illustrated in Figure 10. The figure shows that a model that relies only on mortality risk is unable to account for the observed early claiming. In this model, early claimers are those who lack the assets to afford to delay or find the Social Security annuity unattractive, given their life expectancy. The figure shows that adding out-of-pocket medical expenses does not incentivize early claiming, as these expenses typically occur in later life. The model performs best when bequest motives and health-dependent preferences are incorporated. Bequest motives impact claiming decisions by generating a tradeoff between having more resources in the future and giving up on savings to finance incidental bequests due to market incompleteness. The addition of health-dependent preferences further improves the model's performance because of its effect on impatience and the marginal rate of substitution between consumption and bequests. The subsequent sections provide additional details on comparing the standard model and various extensions with the empirical claiming distribution. During these exercises, we keep the remaining parameters constant and only adjust the ones related to the counterfactual of interest.

Figure 10 Counterfactual Analysis



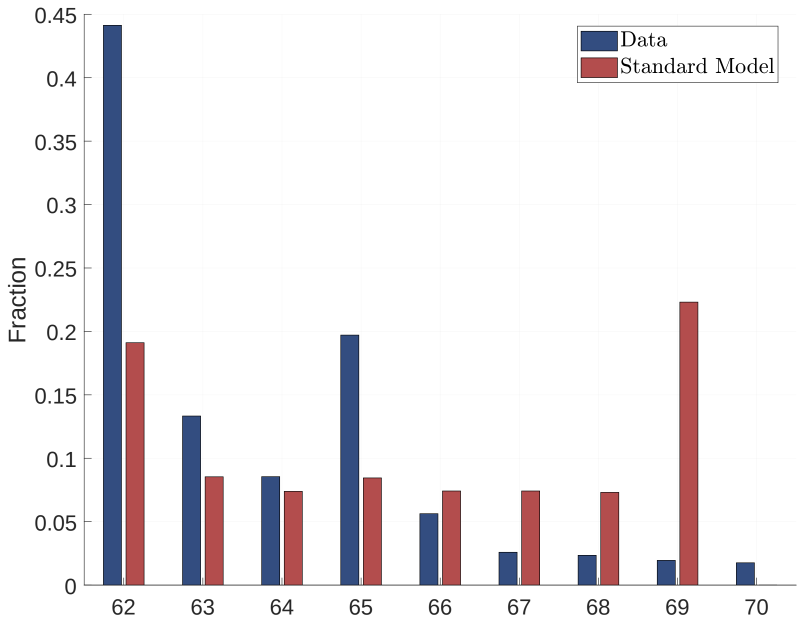
6.1.1 The Standard Model

We define the standard model as our baseline model without bequest motives, health-dependent preferences, and out-of-pocket medical expenses[11]. The goal of this exercise is to establish a baseline for understanding the role of each extension in the baseline model. In the standard model, individuals only face mortality risks. In that sense, individuals maximize utility by choosing the most appealing consumption sequence. If individuals can afford to delay claiming and expect to live long enough, having more resources intertemporally can allow individuals to afford better consumption sequences. Because there is initial heterogeneity in health and, therefore, in life expectancy, there is dispersion in the claiming behavior. This dispersion is also affected by the initial heterogeneity in wealth and income. Figure 11 shows the predicted claiming behavior in this setting. According to the results, a standard life-cycle model that only takes mortality risk into account can only produce 39% of early claimers and 19% of claimers at

[11]The omission of medical expenses includes shutting down both its deterministic and stochastic components.

62. Looking at the results from a different perspective, we can say that 39% of claimers' behavior is not puzzling when analyzed using a standard framework. This is because these individuals' choices can be explained by liquidity constraints and/or low life expectancy, which are featured in the standard model.

Figure 11 Standard Model

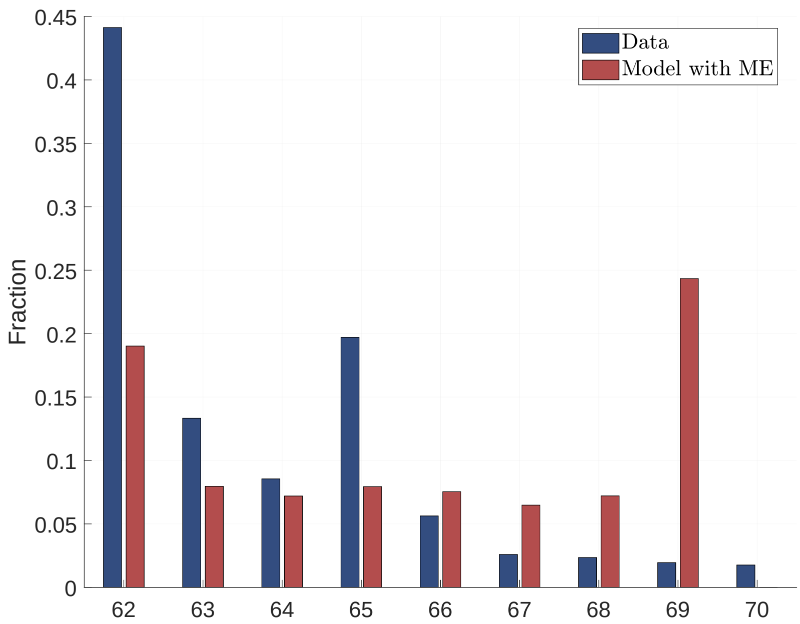


Note: The figure compares the empirical claiming distribution of individuals born before 1937 and the standard model.

6.1.2 Adding Medical Expenses

We next turn to assess the role of medical expenses by adding them to the standard model. Medical expenses may impact claiming decisions by rendering the option of delaying unaffordable, or by affecting the ability of individuals to smooth consumption. These effects, however, depend heavily on the timing of medical expenses. Figure 7 shows that they are concentrated late in life. Because of their timing, medical expenses do not create incentives to claim early. Instead, they amplify the longevity risk, and therefore, its effect goes in the opposite direction. Furthermore, the effect of medical expenses on making delaying an inaccessible option is dampened by the presence of the consumption floor. For instance, poor individuals who could be exposed to the risk of a large medical expense even during the first period have fewer incentives to claim early because the government-provided consumption floor partially covers them. Consistent with this story, Figure 12 shows that adding out-of-pocket medical expenses to the standard framework has a negligible effect in accounting for early claiming.

Figure 12 Standard Model with Medical Expenses

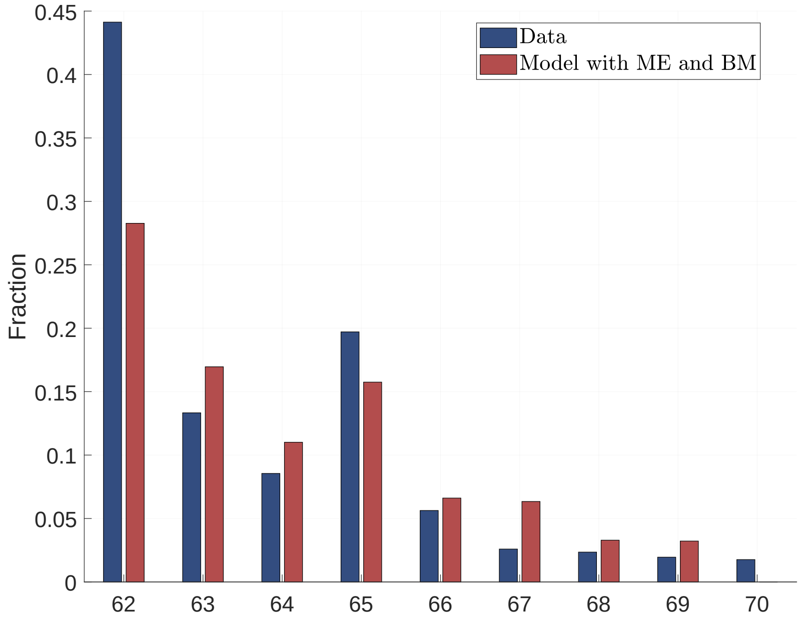


Note: The figure compares the empirical claiming distribution of individuals born before 1937 and the standard model augmented with out-of-pocket medical expenses.

6.1.3 Adding Bequest Motives

Incorporating bequest motives into a standard model that accounts for medical expense risk substantially increases the share of individuals who claim benefits early. As shown in Figure 13, introducing bequest motives raises the proportion of early claimers to nearly 60%, yet it reproduces only 29% of claims at age 62. Because the model features mortality risk, market incompleteness, and borrowing constraints, adding bequest motives activates the mechanism described in our theoretical section: each unit of annuitization reduces the potential for leaving incidental bequests. Moreover, since preferences over bequests are concave, delaying retirement benefit collection entails a trade-off between securing greater terminal resources and maintaining the ability to smooth consumption over the life cycle. It is important to mention that one could potentially fully account for the observed early claiming behavior by increasing the strength of bequest motives. However, this approach has two drawbacks. Firstly, it results in an overaccumulation of assets throughout the life cycle, which adversely affects the model's ability to make reliable counterfactuals. Secondly, even if early claiming is adequately accounted for by bequest motives alone, the model fails to account for consumption fluctuations after retirement caused by health shocks, as explained in the following section.

Figure 13 Standard Model with Medical Expenses and Bequest Motives



Note: The figure compares the empirical claiming distribution of individuals born before 1937 and the standard model augmented with out-of-pocket medical expenses and bequest motives.

6.1.4 Adding Health-dependent preferences

Finally, we add health-dependent preferences to the model, which leads us to our baseline model (Figure 9). This addition increases the fraction of claimers at 62 from 28% to 36% and the fraction of early claimers from 60% to 66%. This increase happens because health follows a downward trend, and therefore, having negative health-dependent preferences reduces the continuation value, making individuals more impatient. Furthermore, because of the downward health trend, health-dependent preferences enhance the effect that bequests have on claiming decisions. In particular, it worsens the ability to smooth over assets given that individuals wish to consume less in the future. In other words, negative health-dependent preferences alter the marginal rate of substitution between consumption and bequests, disincentivizing the delay of the collection of benefits. It is important to notice that for health-dependent preferences to have all their effects, they need the presence of bequest motives. This is because, in the absence of bequests, health dependence acts purely as an amplifier of the discount factor. With only this feature, the model misses three important aspects. First, by only making individuals more impatient, individuals run out of assets too quickly, and therefore, it is not possible to capture the asset-profile accumulation. Second, health dependence amplifies the reluctance to delay that comes from the concavity of the bequest function. Therefore, adding health-dependent preferences without bequest motives leads to underestimating the role of health and its dynamics in accounting for early claiming. Finally, adding health preferences helps to capture the response of consumption to transitory frailty shocks. As shown by Blundell et al. (2020), the consumption fluctuations that happen late in life are accounted for mostly by the negative effect of health on the marginal utility of consumption. Consistent with this result, when health dependence is shut down in our model, the pass-through coefficient to transitory frailty shock goes down to only 0.01. In summary, a model that incorporates health-dependent preferences and bequest motives is consistent with the observed early claiming, the slow deaccumulation of assets after retirement, and the fluctuations of consumption that are due to health and income shocks.

7 Conclusions

This paper extends the standard life-cycle model to account for the observed early claiming behavior of SS benefits. To do it, I depart from documenting a high sensitivity of claiming behavior to changes in health. Using a student event approach, I find that an adverse health event increases the likelihood of claiming by 29%. Based on this fact, I build a quantitative model with a rich set of channels through which health can affect claiming decisions. My framework features a novel channel through which health can impact claiming decisions: the marginal utility of consumption. With negative health-dependent preferences, individuals become more impatient given the downward trend of health and strengthen their bequest motives given the lower value of the marginal utility of consumption. In this way, negative health-dependent preferences reduce the gains from delaying the claiming decision. The calibrated model can successfully account for early claiming. Importantly, counterfactual experiments show that health-dependent preferences and bequest motives are key to this result.

Based on my findings, policymakers who want to ensure individuals against longevity risk should consider health risks when creating policies. This is because there is a connection between individuals' incentives to insure themselves against health and longevity. In future research, it would be helpful to evaluate whether the Social Security system efficiently allocates resources to insure against health and longevity or if there could be potential improvements from redistributing this government-provided insurance.

References

David Altig, Laurence J Kotlikoff, and Victor Yifan Ye. How much lifetime social security benefits are americans leaving on the table? *Tax Policy and the Economy*, 37(1):135–173, 2023.

Richard Blundell, Luigi Pistaferri, and Ian Preston. Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921, 2008.

Richard Blundell, Margherita Borella, Jeanne Commault, and Mariacristina De Nardi. Old age risks, consumption, and insurance. Working Paper 27348, National Bureau of Economic Research, June 2020.

Mariacristina De Nardi. Wealth inequality and intergenerational links. *The Review of Economic Studies*, 71(3):743–768, 2004.

Mariacristina De Nardi, Eric French, and John B Jones. Why do the elderly save? the role of medical expenses. *Journal of political economy*, 118(1):39–75, 2010.

Pierre-Olivier Gourinchas and Jonathan A Parker. Consumption over the life cycle. *Econometrica*, 70(1):47–89, 2002.

Roozbeh Hosseini, Karen A Kopecky, and Kai Zhao. The evolution of health over the life cycle. *Review of Economic Dynamics*, 45:237–263, 2022.

Greg Kaplan and Giovanni L Violante. How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics*, 2(4):53–87, 2010.

Lee M Lockwood. Incidental bequests and the choice to self-insure late-life risks. *American Economic Review*, 108(9):2513–2550, 2018.

Svetlana Pashchenko and Ponpoje Porapakkarm. Accounting for social security claiming behavior. *CRR WP*, (2018-8), 2022.

Nicolo Russo. Health-depdenent preferences, consumption, and insurance. Mimeo, November 2022. URL https://nicolorusso.github.io/jmp_current.pdf.

Menahem E Yaari. Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, 32(2):137–150, 1965.

APPENDIX

A The Effect of health-dependence on Annuitization Decisions

In this sub-section, I show that health dependence affects annuitization decisions. To simplify the problem, we will not consider Social Security income and the decision to claim benefits. I consider an individual who can live for two periods, and their survival probability, given that they are alive in the first period, will I denote by P. The retiree derives utility not only from consumption but also from bequest motives. The utility functions for consumption and bequests are $u(c; f)$ and $\phi(b)$, respectively. It is important to note that the utility function of consumption depends on an additional term, f, which represents frailty. To focus on the effect of health dependence on preferences, we assume that the survival probability is independent of frailty. The individual can transfer resources from the first to the second period using annuities or bonds. Annuities and bonds are purchased by paying units of consumption. In exchange for that, these assets promise a return of R_a (if alive in the second period) and R_b, respectively. Actuarially fair annuities provide $R_a = \frac{R_b}{P}$. In the notation of the main text, $R_a = \frac{1}{p_a}$ and $R_b = \frac{1}{p_b}$. I use this new notation for easier interpretation. Through this section, I will assume that (**A**) holds.

We can write this problem as choosing consumption in the first period, the total savings, and the composition of those savings as follows:

$$\max_{c,A,\omega} \; h(f_1)u(c) + \beta \left[PV\left(RA; f_2\right) + (1-P)\phi\left((1-\omega)R_b A\right)\right],$$

$$s.t.$$

$$c + A = W, \text{ and } R = R_b + \omega\Delta.$$

where $\Delta \equiv R_a - R_b$ and ω denote the fraction of savings composed of annuities. From the first-order conditions, we can show that in an interior solution:

$$h(f_1)u'(c) = \beta\left[PRV'(RA; f_2) + (1-P)(1-\omega)R_b\phi'((1-\omega)R_b A)\right],$$

$$PV'(RA; f_2)\Delta = (1-P)\phi'((1-\omega)R_b A)R_b.$$

So for a given pair of returns R_a and R_b we have that:

$$\frac{\Delta}{R_b} = \frac{(1-P)\phi'((1-\omega)R_b A)}{PV'(RA; f_2)}.$$

In other words:

$$\frac{\Delta}{R_b} = \frac{\text{Expected b's value at 1}}{\text{Expected m's value at 2}}. \tag{35}$$

Equation (35) says that when the continuation value decreases, individuals respond optimally by increasing the share of bonds in their portfolio. The following proposition formalizes this idea:

Proposition 3. *Suppose (A). Then, the optimal annuitization rate responds negatively to higher levels of frailty in period 2, and such a response is given by:*

$$\frac{d\omega}{df_2} = \frac{1}{\sigma}\frac{R(1-\omega)}{\Delta(1-\omega)+R} = \frac{V_{wf}}{V_w}\frac{1}{\sigma}\frac{(R_b+\omega\Delta)(1-\omega)}{R_a} < 0.$$

Also, the optimal annuitization rate responds negatively to an increase in frailty in period 1, and the magnitude depends on the persistence of frailty. In particular:

$$\frac{d\omega}{df_1} = g'(f_1)\frac{\partial\omega}{\partial f_2} < 0.$$

Where $\sigma_g = -\frac{g''(x)}{g'(x)}x$

Proof. From the first-order conditions of the optimization problem we have:

$$p\Delta V_w = (1-p)\phi_b.$$

Taking a derivative with respect to f_2:

$$p\Delta V_{wf} + p\Delta V_{ww}\left[R\frac{dA}{df_2} + A\Delta\frac{d\omega}{df_2}\right] = (1-p)\phi_{bb}\left[(1-\omega)\frac{dA}{df_2} - A\frac{d\omega}{df_2}\right].$$

$$\frac{V_{wf}}{V_w} + \frac{V_{ww}}{V_w}\left[R\frac{dA}{df_2} + A\Delta\frac{d\omega}{df_2}\right] = \frac{\phi_{bb}}{\phi_b}\left[(1-\omega)\frac{dA}{df_2} - A\frac{d\omega}{df_2}\right].$$

$$\left(\frac{V_{ww}}{V_w}\Delta + \frac{\phi_{bb}}{\phi_b}\right)A\frac{d\omega}{df_2} = \left[\frac{\phi_{bb}}{\phi_b}(1-\omega) - \frac{V_{ww}}{V_w}R\right]\frac{dA}{df_2} - \frac{V_{wf}}{V_w}.$$

$$\left(\frac{V_{ww}}{V_w}RA\frac{\Delta}{R} + \frac{\phi_{bb}}{\phi_b}\frac{(1-\omega)A}{1-\omega}\right)\frac{d\omega}{df_2} = -\left(\sigma_\phi - \sigma_v\right)\frac{dA/A}{df_2} - \frac{V_{wf}}{V_w}.$$

$$\left(-\sigma_v \frac{\Delta}{R} - \sigma_\phi \frac{1}{1-\omega}\right) \frac{d\omega}{df_2} = -\left(\sigma_\phi - \sigma_v\right) \frac{dA/A}{df_2} - \frac{V_{wf}}{V_w}.$$

$$\frac{\partial \omega}{\partial f_2} = \frac{\frac{V_{wf}}{V_w} + \left(\sigma_\phi - \sigma_v^{f_2}\right) \frac{dA/A}{df_2}}{\sigma_v^{f_2} \frac{\Delta}{R} + \sigma_\phi \frac{1}{1-\omega}}.$$

Where $\sigma_g = -\frac{g''(x)}{g'(x)} x$. Under Assumption 2, we further have that:

$$\frac{d\omega}{df_2} = \frac{1}{\sigma} \frac{R(1-\omega)}{\Delta(1-\omega) + R} = \frac{V_{wf}}{V_w} \frac{1}{\sigma} \frac{(R_b + \omega\Delta)(1-\omega)}{R_a} < 0.$$

Regarding the effect of a change in frailty in the first period:

$$\frac{d\omega}{df_1} = \frac{\partial \omega}{\partial f} + \frac{\partial \omega}{\partial f_2} \frac{df_2}{df_1} = 0 + g'(f_1) \frac{\partial \omega}{\partial f_2}.$$

□

Proposition 3 states that if frailty reduces the marginal utility of consumption, then a higher level of frailty will decrease the value of consumption, assuming the individual is still alive. In such cases, it is optimal for the individual to switch from annuities to bonds. The proposition also states that whenever there is an increase in frailty during period 1, the optimal annuitization rate will be reduced due to the negative effect of the shock. The extent of this effect will depend on how persistent the frailty is.

B The effect of health-dependent preferences on claiming decisions in a portfolio choice model

In this section, relative to my complete markets setting, I add no-short position constraints:

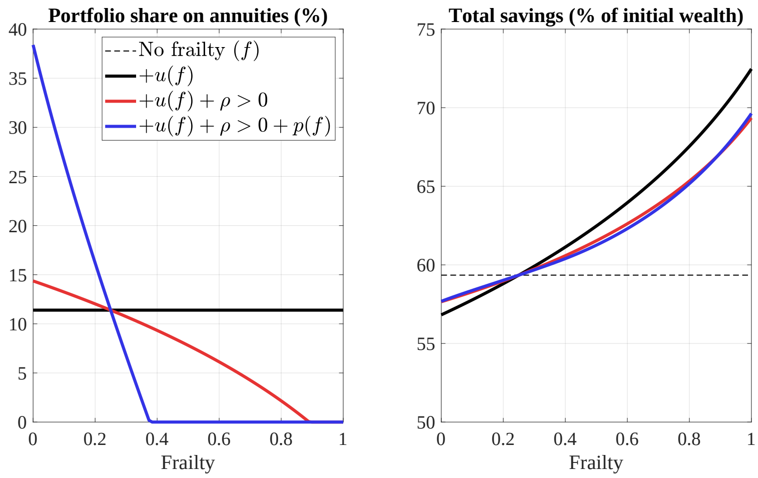
$$v(w, y; \boldsymbol{f_1}, \boldsymbol{f_2}) = \max_{\{c_t\}_t,\, a,\, b,\, d} \quad u(c_1, f_1) + \beta P \left[u(c_2, f_2) + \beta\phi(b_2)\right] + \beta(1-P)\phi(b_1),$$

$$s.t.$$

$$c_1 = w + y_1(d) - p^a a_1 - p^b b_1, \qquad c_2 = a_1 + b_1 + (1 + R_{ss} d) y_2(d) - p^a a_2 - p^b b_2,$$

$$a_1 \geq 0, \quad a_2 \geq 0, \quad b_1 \geq 0, \quad b_2 \geq 0, \quad c_1 \geq 0, \quad c_2 \geq 0.$$

Figure 14 : The impact of frailty-dependent preferences on annuitization decisions.



Note: In the first graph on the left, I show the percentage of total savings represented by annuities in the second period. The percentage is displayed as a function of frailty. The second graph on the right shows the fraction of total savings (bonds plus annuities) as a function of frailty in the second period. The black line indicates that frailty in the second period is not related to frailty in the first period. The red line represents the case in which frailty is persistent, and the blue line represents the adding dependence of survival probability on frailty.

In this scenario, an individual can face a situation where they receive more consumption in the second period than they desired, given a high frailty level. This situation happens when the amount the individual receives from Social Security and bonds exceeds their optimal level of consumption in period two. In a complete market setting, this situation can be resolved by holding negative amounts of annuities. However, due to non-negativity constraints, this is no longer possible. Delaying benefits provides gains in resources for the second period, making it more likely for an individual to face this situation of excess consumption in period two when they opt for delay instead of claiming early. Consequently, the only way to reduce this excess consumption is by reducing bond holdings, which comes at the cost of fewer bequests. If the losses from bequests are higher than the gains in future consumption in terms of utility, then delaying the collection of benefits is no longer optimal. The main result is that under our previous assumptions, there exists a threshold of frailty such that for any frailty value above such threshold, individuals find it optimal to claim early.

Delay: $d = 1$ As the individual chooses to postpone collecting Social Security benefits, she receives annuity income from the government for $y + R_d$, in addition to the income from bonds.

If this total income is insufficient to finance the desired consumption in period 2:

$$\underbrace{c_2^*}_{\text{Future consumption}} \geq \underbrace{(1+R_{ss}d)y}_{\text{SS annuity}} + b.$$

Then, the individual buys annuities with the rest of her wealth to achieve this optimal consumption:

$$a^* = c_2^* - (1+R_{ss}d)y - b. \tag{36}$$

However, when the desired consumption in period 2 when alive is lower than this total income, for instance, because the marginal utility of consumption in the second period is low enough because of the effect of frailty:

$$\underbrace{c_2^* < (1+R_{ss}d)y + b}_{\text{More annuitization than desired}}.$$

The demand for private annuities becomes zero.

Claim early: $d = 0$ If an individual decides to claim early Social Security benefits, there is one advantage: it will be easier for them to achieve their desired level of future consumption. In this case, the annuity income from Social Security will be just y. Specifically, if the demand for consumption in period 2 is higher than what Social Security income and bonds provide ($c_2^* \geq \underbrace{y}_{\text{SS annuity}} + b$), the individual can achieve their optimal consumption by buying:

$$a^* = c_2^* - y - b.$$

This, however, happens at the expense of giving up resources intertemporally (remember, this budget constraint is less generous in the PDV sense). When the desired consumption in period 2 is even lower than what income from bonds and Social Security provides ($c_2^{0,*} < y + b$), the optimal allocation after claiming features $a^{0,*} = 0$.

Thus, whenever individuals have a high demand for annuities, it becomes optimal for them to delay their decision. This is because they can purchase annuities in the private market in addition to what is provided by Social Security, allowing them to choose the desired amount. By doing this, they can expand their lifetime budget constraint. However, when the demand for consumption in the second period is low, the optimal decision is not obvious anymore. The optimal claiming decision, in this case, depends on a critical trade-off between better insurance and higher resources.

B.0.1 The main trade-off: more resources vs. worse insurance

As we saw in the section above, whenever the individual has a lower demand for future consumption than what Social Security together with bond holdings provide, the decision to claim is not trivial anymore because the individual faces a tradeoff between more resources coming from the return or delaying the collection of Social Security benefits and the possibility of achieving the optimal amount of c_2, conditional on surviving, because not receiving the extra income from delaying the collection of benefits, allows the individual to adjust her portfolio better to get closer to the unrestricted demand for c_2. We are interested in the scenario in which $c_2^* < (1 + R_{ss}d)y + b$. In this situation, the individual obtains more annuitization than desired if he decides to delay the collection of benefits. The individual can either delay and have more resources in the lifetime perspective but have an excess of consumption in period 2, or he can claim early and buy annuities until his demand is satisfied but with fewer intertemporal resources. Thus, the decision rule is choosing $d = 1$ if and only if:

$$\underbrace{\beta P\left[u(y(1+R_{ss})+b),\mathbf{f_2}) - u\left(c_2^*,\mathbf{f_2}\right)\right]}_{\text{Gains if alive in second period}} \geq \underbrace{u(\widetilde{C}, f_1) - u(\widehat{c}, f_1)}_{\text{Loses in consumption in period 1}}$$

$$\underbrace{+\beta(1-P)\left[\phi\left(\phi'^{-1}\left(\frac{1}{\beta(1-P)}[p_1-p_0]u'(\widetilde{C},f_1)\right)\right) - \phi\left(\phi'^{-1}\left(\frac{1}{\beta(1-P)}[p_1-p_0]u'(\widehat{c},f_1)\right)\right)\right]}_{\text{Loses in bequests}}. \tag{37}$$

Equation 39 tells us that the decision to delay benefits is optimal if the gains in terms of utility from having a higher income in the second period conditional on surviving outweigh the losses from enjoying less consumption in the first period and the losses from leaving fewer bequests. Herefore, if the marginal utility of consumption in the second period is low enough (because of frailty), then delaying the collection of benefits is not optimal anymore. The following proposition formalizes this idea.

Proposition 4. *Under (A), there exists a* $f^* > 0$ *such that for any* $f_2 \leq f^*$, $d(f_2) = 1$, *while for* *any* $f_2 > f^*$, $d(f_2) = 0$

Proof. From the optimization problem, we get the following first-order conditions:

$$c_2^0 \leq (u')^{-1}\left(\frac{p^0 u'(c_1, f_1)}{P\beta}, f_2\right),$$

$$c_2^1 \leq (\phi')^{-1}\left(\frac{1}{(1-P)\beta}\left[p_1 - p_0\right] u'(c_1, f_1)\right),$$

with strict equality if the non-negativity constraints are nonbinding. In the latter case, the intertemporal budget constraint is:

$$c_1 + p^0 (u')^{-1}\left(\frac{p^0 u'(c_1, f_1)}{P\beta}, f_2\right) + (p^1 - p^0)(\phi')^{-1}\left(\frac{1}{\beta(1-P)}\left[p_1 - p_0\right] u'(c_1, f_1)\right)$$

$$= w + y + p^0(R_d - p_d)d.$$

Denote by $\widehat{C}$ as the quantity that solves a relaxed problem (non-binding short position constraints):

$$\widehat{C} + p^0 (u')^{-1}\left(\frac{p^0 u'(\widehat{C}, f_1)}{P\beta}, f_2\right) + (p^1 - p^0)(\phi')^{-1}\left(\frac{1}{\beta(1-P)}\left[p_1 - p_0\right] u'(\widehat{C}, f_1)\right)$$

$$= w + y + p^0(R_d - p_d)d.$$

For each possible decision, let's analyze what the demand for annuities looks like:

Delay: $d = 1$ When the individual decides to delay the collection of Social Security benefits, she receives annuity income from the government for an amount of $y + R_d$ and income from bonds. If this total income is insufficient to finance the desired consumption in period 2:

$$p^0 \underbrace{(u')^{-1}\left(\frac{p^0 u'(\widehat{C}, f_1)}{\beta(1-m)}, f_2\right)}_{\text{annuity demand}} \geq \underbrace{y + R_d}_{\text{SS annuity}} + b.$$

Then, the individual annuitizes the rest of her wealth to achieve this optimal amount. These annuity purchases are illustrated in the following equation:

$$a^0 = p^0 (u')^{-1}\left(\frac{p^0 u'(\widehat{C}, f_1)}{\beta(1-m)}, f_2\right) - y - R_d - b. \tag{38}$$

where $c_1 = \widehat{C}$. However, when the desired consumption in period 2 is lower than this total income:

$$\underbrace{p^0 (u')^{-1}\left(\frac{p^0 u'(\widehat{C}, f_1)}{\beta(1-m)}, f_2\right) < y + R_d + b}_{\text{More annuitization than desired}}.$$

The demand for private annuities becomes zero, and the consumption in the first period is $c_1 = \hat{c}$, where $\hat{c}$ is the quantity that solves:

$$\widehat{c} + (p^1 - p^0)(\phi')^{-1}\left(\frac{1}{\beta(1-P)}\left[p_1 - p_0\right] u'(\widehat{c}, f_1)\right) = w + y - p^0 p_d.$$

Claim early: $d = 0$ If the individual decides to claim early and the demand for consumption in period 2 is higher than what Social Security income and bonds provide ($p^0 (u')^{-1}\left(\frac{p^0 u'(\widehat{C}, f_1)}{\beta(1-m)}, f_2\right) \geq y + b$), the individual can achieve her optimal consumption by buying:

$$a^0 = (u')^{-1}\left(\frac{p^0 u'\left(\widetilde{C}, f_1\right)}{\beta(1-m)}, f_2\right) - y - b.$$

where $c_1 = \widetilde{C}$ and $\widetilde{C}$ is the quantity that solves:

$$\widetilde{C} + p^0 (u')^{-1}\left(\frac{p^0 u'(\widetilde{C}, f_1)}{P\beta}, f_2\right) + (p^1 - p^0)(\phi')^{-1}\left(\frac{1}{\beta(1-P)}\left[p_1 - p_0\right] u'(\widetilde{C}, f_1)\right)$$
$$= w + y.$$

This, however, happens at the expense of giving up resources intertemporally since $\widehat{C} > \widetilde{C}$. When the desired consumption in period 2 is even lower than what income from bonds and Social Security provides ($(u')^{-1}\left(\frac{p^0 u'\left(\widetilde{C}, f_1\right)}{\beta(1-m)}, f_2\right) < y + b$), the optimal allocation after claiming is given by $a^0 = 0$ and $c_1 = \widetilde{c}$ where $\widetilde{c}$ is the quantity that solves:

$$\widetilde{c} + (p^1 - p^0)(\phi')^{-1}\left(\frac{1}{\beta(1-P)}\left[p_1 - p_0\right] u'(\widetilde{c}, f_1)\right) = w + y.$$

We are interested in the scenario in which the demand for consumption in the second period:

$y + b \leq c_2^0 < y + R_d + b$. If the individual decides to delay, the intertemporal utility is given by:

$$U_{delay} = u(\widehat{c}, f_1) + \beta \left[Pu(y + R_d + b, f_2) + (1 - P)\phi \left(\phi'^{-1} \left(\frac{1}{\beta(1-P)} [p_1 - p_0] u'(\widehat{c}, f_1) \right) \right) \right].$$

The utility from claiming in the first period is given by:

$$U_{claim} = u(\widetilde{C}, f_1)$$

$$+\beta \left[Pu \left((u')^{-1} \left(\frac{p^0 u' \left(\widetilde{C}, f_1 \right)}{\beta(1-m)}, f_2 \right), f_2 \right) + (1 - P)\phi \left(\phi'^{-1} \left(\frac{1}{\beta(1-P)} [p_1 - p_0] u'(\widetilde{C}, f_1) \right) \right) \right].$$

Thus, the decision rule is choosing $d = 1$ if and only if:

$$\underbrace{\beta P \left[u(y + R_d + b, f_2) - u \left(c_2^{0,*}, f_2 \right) \right]}_{\text{Gains if alive in second period}} \geq \underbrace{u(\widetilde{C}, f_1) - u(\widehat{c}, f_1)}_{\text{Loses in consumption in period 1}}$$

$$\underbrace{+\beta(1 - P) \left[\phi \left(\phi'^{-1} \left(\frac{1}{\beta(1-P)} [p_1 - p_0] u'(\widetilde{C}, f_1) \right) \right) - \phi \left(\phi'^{-1} \left(\frac{1}{\beta(1-P)} [p_1 - p_0] u'(\widehat{c}, f_1) \right) \right) \right]}_{\text{Loses in bequests}} \tag{39}$$

This can be written as:

$$\beta P u'(c_2^{0,*}, f_2)(y + R_d + b - c_2^{0,*}) \geq u'(\widehat{c}, f_1)(\widetilde{C} - \widehat{c}) + \beta(1 - P)\phi'(b_{delay}) \left(b_{claim} - b_{delay} \right).$$

Therefore, for a marginal utility in period 2 that is low enough, the condition above does not hold, proving the proposition.

□

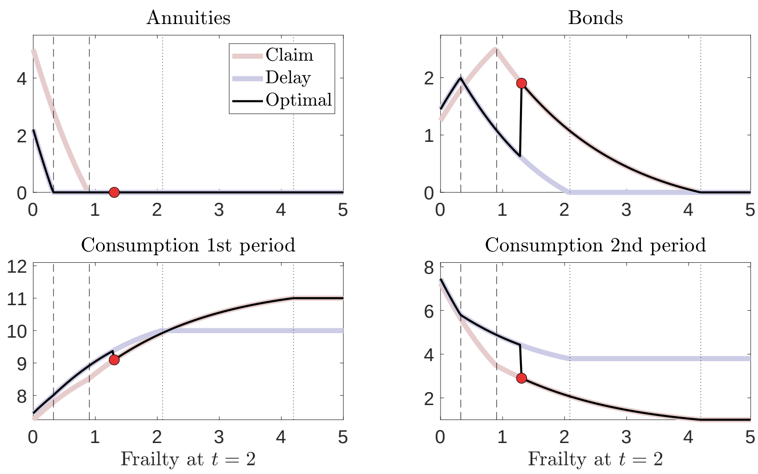
The intuition is as follows: when the value of consumption of period 2 has a sufficiently low value for the individual (for instance, because bad health conditions make consuming in period two less appealing), then it becomes less likely for the individual to give up on consumption in period one and bequests to have a higher annuitization level. In particular, when frailty negatively affects the marginal utility of consumption, it affects the claiming decision through two channels. First, when frailty in period two is higher than in period one, for a given level of consumption, the marginal utility of consumption in period two is lower than in period 1. Second, because the marginal utility of bequests does not depend on health in period two, we have that when frailty affects the marginal utility of consumption in period two if alive, the marginal rate of substitution between bequests and consumption changes. In particular, bequests become more appealing relative to consuming sick.

The mechanism described in Figure 15 shows the optimal quantities of consumption and savings for individuals based on their level of frailty, depending on whether they claim benefits in the first period or delay. The red line in the figure represents the optimal consumption and savings for those who claim benefits in the first period, while the blue lines represent those who choose to delay. The demand for consumption in period two decreases as frailty increases. However, there is a point at which individuals reduce their consumption in period two, which happens when the non-negativity constraint for annuities becomes binding. This point occurs at a lower level of frailty for those who decide to delay, as they have higher consumption in period two and, therefore, require fewer annuities. When the non-negativity constraints are binding, individuals should delay the collection of benefits, as it results in more consumption in both periods and more savings in bonds to leave bequests. However, when the non-negativity constraint for annuities becomes binding for those who decide to delay, there is a tradeoff. The excess consumption in period two can only be reduced by decreasing savings in bonds, which results in lower bequests than if they had claimed benefits in the first period. The red dot in the plot indicates the level of frailty at which the losses from delaying surpass the gains from delaying. This is the point at which the optimal decision is to claim in the first period. The bottom panel of Figure 15 shows how the frailty level corresponding to the red dot is the threshold at which claiming in the first period becomes the optimal decision. This is the point at which the losses from delaying become more significant than the gains from delaying.

Overall, this analysis suggests that in a framework with non-negativity constraints, the individual might decide to claim benefits early to insure against health or at least mitigate the effects of bad health. A way of doing this would be by issuing negative amounts of annuities or borrowing if the effect of frailty in the marginal utility in the second period is sufficiently high. However, these constraints can leave an individual with no choice but to claim early.

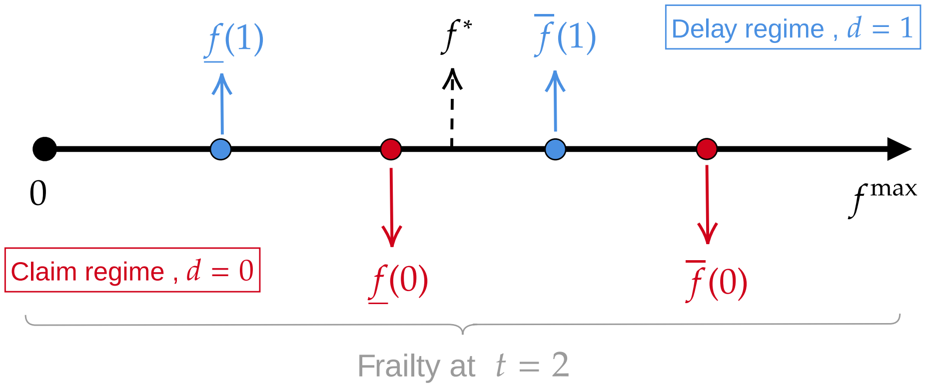
Something worth mentioning here is that in this example and through this paper, the non-negativity constraint that binds first is the one regarding annuities rather than the one for bonds, regardless of the claiming decision. This happens for two reasons: first, because I assume that annuities are cheaper than bonds, and also because bonds give utility through bequest motives. This can be illustrated in Figure 16. In this figure, I illustrate the cutoffs of frailty at which the non-negativity constraints become binding for each claiming choice. $\underline{f}(1)$ and $\overline{f}(1)$ denote the frailty levels at which, conditional on choosing to delay, the non-negativity constraints for annuities and bonds become binding, respectively. $\underline{f}(0)$ and $\overline{f}(0)$ denote the same for an individual who chooses to claim benefits in the first period. Since, in this case, buying annuities to finance consumption in period two is cheaper than buying bonds for this purpose, we have that to reduce consumption in the second period, it is better to reduce annuitization first rather than savings in bonds. This is why $\underline{f}(1) < \overline{f}(1)$ and $\underline{f}(0) < \overline{f}(0)$. Because delaying implies more consumption in the second period, we have that $\underline{f}(1) < \underline{f}(0)$ and $\overline{f}(1) < \overline{f}(0)$. In this plot, f^* denotes the threshold at which claiming in the first period becomes optimal. A feature of f^* is that $f^* > \underline{f}(1)$. This is because, for lower values of frailty, we are in the unrestricted case (complete markets), in which the optimal decision consists of maximizing the Present Discounted Value of benefits.

Figure 15 : Optimal quantities as a function of frailty in period 2.



Note: Red dot represents the threshold of frailty in period 2 at which claiming early becomes optimal.

Figure 16 : The frailty cutoff of claiming decisions.



Note: Frailty threshold for the shift in the optimal claiming decision. $\underline{f}(0)$ and $\underline{f}(1)$ denote the thresholds of frailty at which the non-negativity constraint for annuities becomes binding. $\overline{f}(0)$ and $\overline{(f)}(1)$ denote the frailty thresholds at which the non-negativity constraint in bonds becomes binding. f^* denotes the threshold frailty at which for any frailty in the second period higher than f^* the optimal decision consists of claiming in period 1.

C Proofs

C.1 Proof of Proposition 1

From the first-order conditions of this problem, we get:

$$u'(c_1(d), f_1) = \frac{\beta}{p^b} \left[(1-P)\phi'(b_1(d)) + R_b P u'(c_2(d), f_2)\right], \tag{40}$$

$$u'(c_2(d), f_2) = \frac{p^a u'(c_1(d), f_1)}{\beta P}, \tag{41}$$

and

$$u'(c_2(d), f_2) = \frac{\beta \phi'(b_2(d))}{p^b}. \tag{42}$$

From 41 and 42, we can see that both $c_2(d)$ and $b_2(d)$ can be expressed as a function of $c_1(d)$. Furthermore, given that $c_2(d)$ can be expressed as a function of $c_1(d)$. From 40, we can show that $b_1(d)$ can also be expressed as a function of $c_1(d)$. From the budget constraint of the first period, we can show that $a_1(d)$ can be expressed as a function of $c_1(d)$ because all the other variables are already functions of $c_1(d)$. Finally, because $a_2(d) = 0$ since the individual only lives up to 2 periods, we can combine the budget constraints to obtain an intertemporal budget constraint:

$$c_1 + p^a(c_2(c_1) - b_1(c_1) - y_2(d) + p^b b_2(c_1)) + p^b(b_1(c_1)) = w + y_1(d).$$

Which can be expressed as:

$$c_1 + p^a(c_2(c_1)) + (p^b - p^a)b_1(c_1) + p^a p^b b_2(c_1) = w + y_1(d) + p^a y_2(d).$$

Since all variables are increasing functions of c_1, it follows that the optimal claiming decision consists of maximizing $y_1(d) + p^a y_2(d)$, which, in the case we are interested in, consists of delaying the collection of benefits ($d = 1$).

C.2 Proof of Proposition 2

We can rewrite this problem as follows:

$$V(f_2, d) = \max_{b_1} u(y_1(d) - b_1, f_1) + \beta P[V_2(Rb_1, f_2, d)] + \beta(1 - P)\phi(b_1),$$

$$s.t.$$

$$c_1 + b_1 = y_1(d),$$

From this problem, we can obtain the following first-order condition:

$$u'(y_1(d) - b_1, f_1) = \beta\left[PV'_{2,b_1}(Rb_1, f_2, d) + (1 - P)\phi'(b_1)\right],$$

The first step to prove this proposition is showing that worse health in the second period reduces consumption in the second period, given a claiming choice. To do this, first notice that we can write V_2 as a terminal problem:

$$V_2(Rb_1, f_2, d) = \underset{b_2}{Max} \quad u(Rb_1 + y_2(d) - b_2, f_2) + \beta\phi(b_2),$$

From the envelope theorem, we have that:

$$V'_{2,b_1} = Ru'(c_2, f_2)\Big|_{c=Rb_1+y_2(d)-b_2} > 0,$$

$$V'_{2,f_2} = \left. \frac{\partial u(c_2, f_2)}{\partial f_2} \right|_{c=Rb_1+y_2(d)-b_2} < 0.$$

The following lemma shows that when frailty in the second period negatively affects the marginal utility of consumption, then the individual shifts resources in the second period towards more bequests.

Lemma 1. *If u and ϕ satisfy the usual assumptions , $\frac{\partial^2 u(c,f_2)}{\partial f_2 dc} < 0$, then c_2 is decreasing in f_2*

Proof. It is enough to see that from the first-order conditions of the terminal problem, we have:

$$u'(c_2, f_2) = \beta\phi'(Rb_1 + y_2(d) - c_2).$$

From the concavity of the utility functions and our assumption of the effect of frailty on the marginal utility of consumption, we have that for this equality to hold for higher levels of frailty, c_2 must be decreasing in f_2. This, of course, implies that b_2 is increasing in f_2. □

With this information, we can show that frailty reduces the incentives to save in the first period. In other words, it induces impatience:

Lemma 2. *If $u'(c, \cdot)$ is decreasing $\forall c$ and $\frac{\partial^2 u(c,f_2)}{\partial f_2 dc} < 0$, $b_1(\cdot, d)$ is non-increasing for any $d \in \{0, 1\}$.*

Proof.

$$G(b_1, d) \equiv u'(y_1(d) - b_1, f_1) - \beta(1-P)\phi'(b_1) = \beta P V'_{2,b_1} = \beta P R u'(c_2, f_2).$$

Let $b_1 = b_1(f_2, d)$ and $b'_1 = b_1(f'_2, d)$ with $f'_2 > f_2$. Then we have:

$$G(b_1, d) = u'(y_1(d) - b_1, f_1) - \beta(1-P)\phi'(b_1) = \beta P V'_{2,b_1} = \beta P R u'(c_2, f_2).$$

$$G(b'_1, d) = u'(y_1(d) - b'_1, f_1) - \beta(1-P)\phi'(b'_1) = \beta P V'_{2,b'_1} = \beta P R u'(c_2, f'_2).$$

As a matter of a contradiction, suppose $b'_1 > b_1$. Then we have:

$$G(b_1, d) - G(b'_1, d) = \beta P R[u'(Rb_1 + y_2(d) - b_2, f_2) - u'(Rb'_1 + y_2(d) - b'_2, f'_2)] > 0.$$

However,

$$u'(y_1(d) - b_1, f_1) - u'(y_1(d) - b_1', f_1) - \beta(1 - P)\left[\phi'(b_1) - \phi'(b_1')\right] < 0.$$

Therefore, $b_1(., d)$ is non-increasing in f_2 for any $d \in \{0, 1\}$. □

The following lemma shows that when delaying gives more resources intertemporally at the expense of fewer resources in the first period, at the optimal allocation, there are more resources in the second period:

Lemma 3. *Suppose* $y_1(1) + R^{-1}y_2(1) > y_1(0) + R^{-1}y_2(0)$. *Then, at the optimal allocation:*

$$Rb_1(f_2, 1) + y_2(1) \geq Rb_1(f_2, 0) + y_2(0).$$

Proof. As before, let's define the following terms:

$$G(b_1, 1) = u'(y_1(1) - b_1, f_1) - \beta(1 - P)\phi'(b_1) = \beta PRu'(Rb_1 + y_2(1), f_2),$$

$$G(b_1', 0) = u'(y_1(0) - b_1', f_1) - \beta(1 - P)\phi'(b_1') = \beta PRu'(Rb_1' + y_2(0), f_2).$$

As a matter of a contradiction, suppose:

$$Rb_1(f_2, 1) + y_2(1) < Rb_1(f_2, 0) + y_2(0).$$

Then:

$$G(b_1, 1) - G(b_1', 0) = \beta PR\left[u'(Rb_1 + y_2(1), f_2) - u'(Rb_1' + y_2(0), f_2)\right] > 0.$$

Since $y_1(1) < y_1(0)$ and $y_2(1) > y_2(0)$, then we have:

$$b_1 - b_1' < \frac{1}{R}(y_2(0) - y_2(1)) < y_1(1) - y_1(0) < 0.$$

Which implies:

$$b_1 < b_1'.$$

However, this implies:

$$G(b_1,1) - G(b_1',0) = u'(y_1(1) - b_1, f_1) - u'(y_1(0) - b_1', f_1) - \beta(1-P)\left(\phi'(b_1) - \phi'(b_1')\right) < 0.$$

Which is a contradiction. Therefore:

$$Rb_1(f_2,1) + y_2(1) \geq Rb_1(f_2,0) + y_2(0).$$

□

What follows now is to prove that whenever there are more resources in the second period, there is also more consumption, given a f_2. The following proposition formalizes this idea:

Lemma 4. *If* $\frac{\partial u}{\partial c} > 0$*, then if* $y_1(1) + R^{-1}y_2(1) > y_1(0) + R^{-1}y_2(0)$*,* $c_2(f_2,1) > c_2(f_2,0)$*.*

Proof. From the previous proposition, we know that:

$$Rb_1(f_2,1) + y_2(1) \geq Rb_1(f_2,0) + y_2(0).$$

Define From the first-order conditions:

$$G(b_1(f_2,1),1) = u'(c_2(f_2,1),f_2) = \beta\phi'(Rb_1(f_2,1) + y_2(1) - c_2(1)).$$

$$G(b_1(f_2,0),0) = u'(c_2(f_2,0),f_2) = \beta\phi'(Rb_1(f_2,0) + y_2(0) - c_2(0)).$$

Suppose as a matter of a contradiction that $c_2(f_2,1) \leq c_2(f_2,0)$, then:

$$G(b_1(f_2,1),1) - G(b_1(f_2,0),0) = \beta\left[\phi'(Rb_1(f_2,1) + y_2(1) - c_2(1)) - \phi'(Rb_1(f_2,0) + y_2(0) - c_2(0))\right]$$

$$< 0.$$

However,

$$u'(c_2(f_2,1),f_2) - u'(c_2(f_2,0),f_2) \geq 0.$$

Which is a contradiction. Therefore, $c_2(f_2,1) > c_2(f_2,0)$. □

Finally, we show that when individuals are better off by delaying when they are perfectly healthy, then if the effect of frailty on the marginal utility of consumption is strong enough, there is a threshold such that for higher levels of frailty, the optimal decision is to claim in the first period. Let $c_2(f_2, d) \equiv Rb_1(f_2, d) + y_2(d) - b_2(f_2, d)$. Then, from differentiability:

$$V(f_2, d) = V(0, d) + \int_0^{f_2} \frac{\partial V(d, \widetilde{f_2})}{\partial \widetilde{f_2}} d\widetilde{f_2}.$$

And from the envelope condition:

$$= V(0, d) + \beta P \int_0^{f_2} \frac{\partial u(c_2(\widetilde{f_2}, d), \widetilde{f_2})}{\partial \widetilde{f_2}} d\widetilde{f_2}.$$

Therefore:

$$V(f_2, 1) - V(f_2, 0) = \underbrace{\Delta}_{>0} + \beta P \int_0^{f_2} \left[\frac{\partial u(c_2(\widetilde{f_2}, 1), \widetilde{f_2})}{\partial \widetilde{f_2}} - \frac{\partial u(c_2(\widetilde{f_2}, 0), \widetilde{f_2})}{\partial \widetilde{f_2}} \right] d\widetilde{f_2}.$$

From the previous proposition, we know that $c_2(\widehat{f_2}, 1) > c_2(\widehat{f_2}, 0) \forall \widehat{f_2}$. Therefore,

$$\beta P \int_0^{f_2} \left[\frac{\partial u(c_2(\widetilde{f_2}, 1), \widetilde{f_2})}{\partial \widetilde{f_2}} - \frac{\partial u(c_2(\widetilde{f_2}, 0), \widetilde{f_2})}{\partial \widetilde{f_2}} \right] d\widetilde{f_2} < 0.$$

Thus, if $\left| \frac{\partial^2 u(c, f_2)}{\partial f_2 dc} \right|$ is large enough, the optimal decision consists of claiming in the first period.

D HRS and CAMS data

The Health and Retirement Study (HRS) is a survey that spans over time and is representative of the U.S. population aged 50 and above, along with their spouses. It provides comprehensive information on health, income, assets, individual Social Security decisions, demographics, and other variables. The CAMS questionnaire has been answered by a subset of HRS households every other year since 2001. I solely use the CAMS data to estimate the pass-through coefficients of frailty shocks and earnings shocks. When using the CAMS data, we merge it with the HRS. For every other section, I only use the HRS data. For the HRS data exercises, the sampling procedure is as follows: To construct moments related to claiming behavior, I use the RAND HRS 2018 (V2). We consider respondents born before 1937 who were not receiving Disability Insurance benefits. To estimate the frailty, income, and out-of-pocket medical expenses process, we pool all the HRS waves.

E Variables Definition

E.1 Frailty index

Table 9 lists the 36 variables we used to construct the frailty index for HRS respondents. As in Hosseini et al. (2022), the index is constructed by summing the variables in the first column of the table using their values, which are assigned according to the rules in the second column. Then, divide this sum by the total number of variables observed for the individual in the year. In my framework, I ignore the possibility of frailty taking the value of zero in the next period if it takes the value of zero today. This is because at the age of 62, only 0.51% of individuals in my sample feature that.

E.2 Medical Expenses

Medical expenses include four categories of expenditures: health insurance premia, drugs, health services, and medical supplies. The expenses are in dollars of 2021, which is my base year in this paper. Drugs is defined as the out-of-pocket expenses on prescription and nonprescription medications. Medical services and supplies include out-of-pocket expenses in hospital care, doctor services, lab tests, eye, dental, nursing home care, and medical supplies.

E.3 Consumption

I measure consumption as nondurable consumption, including 21 items: electricity, water, heat, phone and housing supplies and services, garden supplies and services, food, dining out, clothing, vacations, tickets, hobbies, sports equipment, contributions and gifts, personal care, auto insurance, vehicle services, and gasoline. Consumption is deflated by the price index for total consumption of the BLS.

E.4 Income

I measure income as any source of income that is not retirement benefits from Social Security. This includes wages, salaries and bonuses, capital income, self-employment, dividend and interest income, rents, pensions, annuity income, transfers, and other income, including alimony and lump sums from insurance. Income is deflated by the price index for total consumption of the BLS. The household's income is adjusted by family size using the OECD household equivalence scale.

E.5 Assets

Assets refer to the total value of a household's net worth, which is adjusted based on the family size using the OECD household equivalence scale. This includes the sum of all assets, such as the primary and secondary residences, other real estate, vehicles, businesses, Individual Retirement Accounts (IRA), stocks, mutual funds, investment trusts, checking and savings accounts, Certificates of Deposit (CD), government saving bonds, T-bills, bonds, and all other savings, minus all debts, including mortgages, other home loans, among others. To calculate the actual value of assets, we use the price index for total consumption provided by the BLS, and to adjust the value of household assets, we use the OECD household equivalence scale based on the family size.

F Measuring the strength of bequest motives

To measure the strength of bequest motives with other studies in the literature, I follow Pashchenko and Porapakkarm (2022). In particular, to allow for comparability in the strength of bequest motives with other studies, all estimates are converted into two parameters: Marginal Propensity to Bequeath (MPB) and the bequest threshold ($\bar{b}$). To obtain these parameters, consider the simple problem of an individual in its last period of life. The individual has to allocate his available cash on hand x between consumption and bequests. The lifetime utility, denoted by V, is:

$$V = u(c, f) + \beta\phi(b)$$

where u denotes the utility from consumption, which is health dependent, and ϕ is the function governing the utility from bequests. Assuming $\phi(0) < \infty$, which holds under my assumed functional form, the optimal amount of bequests (b^*) can be found from the first-order conditions:

$$u'(x - b^*, f) = \beta\phi'(b^*). \tag{43}$$

Since the marginal utility of bequests is bounded, $b^* = 0$ is a possibility. Then, the bequest threshold $\overline{b}$ is defined as the level of assets such that it is optimal not to leave bequests if $x \leq \overline{b}$. In other words, $\overline{b}$ solves:

$$u'(\overline{b}, f) = \beta\phi'(0). \tag{44}$$

Under my functional specification, this threshold is given by:

$$\overline{b} = \left(\frac{\beta\phi_1}{1 - \delta f}\right)^{-\frac{1}{\sigma}} \phi_2. \tag{45}$$

As can be seen, the higher the value of frailty or the degree of negative health dependence, the lower this threshold, meaning that in the presence of health-dependent preferences, keeping everything constant, there is a larger fraction of individuals leaving bequests. For those who leave bequests, the MPB is defined as follows:

$$MPB = \frac{\partial b^*}{\partial x}. \tag{46}$$

Given my functional specification, this parameter is given by:

$$MPB = \frac{1}{1 + \left(\frac{\phi_1\beta}{1-\delta f}\right)^{-\frac{1}{\sigma}}}. \tag{47}$$

As can be observed, for any parameter values in the bequest utility function, adding negative health dependence increases the marginal propensity to bequeath. The intuition from this comes from the fact that when consumption loses value because of health, then it is optimal to consume less and leave more bequests. A problem with health-dependence, though, is that no measure of the strength of bequest motives entirely depends on the parameters of the bequest function. Instead, these values can vary with frailty.

G Additional specifications - The effect of health expectations on claiming

G.1 Probit Specification

G.2 Logit Specification

H Additional Calibration Details

This section includes information, moments, identification, and estimation results associated with the quantitative model.

H.1 Mortality

Table 12 shows the estimation results of the Probit regression for mortality. Overall, the likelihood of death increases with age and frailty. Because the data is biannual, I approximate the survival rate by taking the square root of the estimates.

H.2 Frailty Process

To control for selection bias due to mortality, I estimate the frailty process using SMM. The estimation process involves defining a set of targeted moments consisting of two sets of moments. The first set of moments comprises the means of the logarithm of frailty in two-year age groups from 50 to 95. The second set of moments captures the stochasticity of frailty. It includes the variance and the first five autocovariance of the components of frailty that cannot be accounted for by observables. To compute these moments, I first extract the variation in frailty due to age and education by regressing frailty to a polynomial on age and education attainment dummies. I then calculate the mean of the square of the residuals by age group. To control for cohort effects, I regress the square of the residuals on cohort dummies and a set of age dummies and remove the estimated cohort effects. Finally, I rescale the cohort-adjusted variances by the raw variance at 60, so the adjusted variance at 60 is the same as the raw variance. To compute the autocovariance profile, I follow a similar process. Given a set of parameters, denoted by θ, and simulating moments denoted by $m^s(\theta)$, I find the parameters by minimizing the loss function, where m^e is the set of empirical moments.

$$\min_{\theta} \quad (m^e - m^s(\theta))')(m^e - m^s(\theta))'),$$

in other words, I will use an identity matrix as a weighting matrix for this optimization. Table 13 shows the estimation results.

H.3 Income Process

The income process is estimated in two steps. First, I regress the income adjusted by family size and by the CPI with a set of covariates that includes age, education, and cohort dummies. Then, I compute the residuals of this regression and compute their variance and autocovariance matrix. Because of the assumed stochastic process, it is possible to identify the parameters governing the persistent and transitory shocks by using an equally weighted minimum distance. I target up to the autocovariance of the 5th order. Table 14 show the estimation results.

H.4 Medical Expenses

The medical expenses process is obtained by estimating the log of out-of-pocket medical expenses with a polynomial in age, a polynomial in frailty, and education. To estimate the variance of the transitory shock, I compute the squared residuals of the previous regression and take its variance as my estimate for the parameter that governs the transitory shock. The results are shown in Table 15.

Table 9 List of health deficits employed to construct frailty

Variable	Value
Some difficulty with ADL/IADLs	
Eating	Yes= 1, No= 0
Dressing	Yes= 1, No= 0
Getting in/out of bed	Yes= 1, No= 0
Using the toilet	Yes= 1, No= 0
Bathing/shower	Yes= 1, No= 0
Walking across room	Yes= 1, No= 0
Walking several blocks	Yes= 1, No= 0
Using the telephone	Yes= 1, No= 0
Managing money	Yes= 1, No= 0
Shopping for groceries	Yes= 1, No= 0
Preparing meals	Yes= 1, No= 0
Getting up from chair	Yes= 1, No= 0
Stooping/kneeling/crouching	Yes= 1, No= 0
Lift/carry 10 lbs	Yes= 1, No= 0
Using a map	Yes= 1, No= 0
Taking medications	Yes= 1, No= 0
Climbing 1 flight of stairs	Yes= 1, No= 0
Picking up a dime	Yes= 1, No= 0
Reaching/extending arms up	Yes= 1, No= 0
Pushing/pulling large objects	Yes= 1, No= 0
Ever had one of the following conditions	
High Blood Pressure	Yes= 1, No= 0
Diabetes	Yes= 1, No= 0
Cancer	Yes= 1, No= 0
Lung disease	Yes= 1, No= 0
Heart disease	Yes= 1, No= 0
Stroke	Yes= 1, No= 0
Psychological problems	Yes= 1, No= 0
Arthritis	Yes= 1, No= 0
BMI $\geq$ 30	Yes= 1, No= 0
Has ever smoked	Yes= 1, No= 0
Back pain	Yes= 1, No= 0
Doctor visit	Yes= 1, No= 0
Hospital visit	Yes= 1, No= 0
Home care visit	Yes= 1, No= 0
Nursing home stay	Yes= 1, No= 0
Cognitive impairment score	0-35, inverted and rescaled to 0-1

Table 10 Claiming likelihood and health expectations (Probit Specification)

	Any	10k	100k	500k
Frailty (Q2)	-0.74	-1.20	-1.56*	-1.46
	(1.23)	(0.95)	(0.93)	(0.94)
Frailty (Q3)	5.88***	5.61***	5.70***	5.72***
	(1.05)	(0.78)	(0.77)	(0.77)
Frailty (Q4)	1.29	0.85	0.42	0.54
	(1.24)	(0.95)	(0.93)	(0.93)
Frailty (Q1) × Frailty expectations	0.06***	0.06***	0.06***	0.07***
	(0.02)	(0.01)	(0.01)	(0.01)
Frailty (Q2) × Frailty expectations	0.08***	0.09***	0.10***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)
Frailty (Q3) × Frailty expectations	-0.08***	-0.07***	-0.08***	-0.07***
	(0.02)	(0.01)	(0.01)	(0.01)
Frailty (Q4) × Frailty expectations	0.03*	0.04***	0.04***	0.04***
	(0.02)	(0.01)	(0.01)	(0.01)
Sex	0.08	-0.01	0.00	0.01
	(0.10)	(0.08)	(0.08)	(0.08)
Liquid Wealth	-0.11*	-0.06***	-0.06***	-0.04*
	(0.06)	(0.02)	(0.02)	(0.02)
Mortality risk	-0.82	-0.11	-0.15	-0.23
	(1.29)	(1.03)	(1.05)	(1.04)
Has college	(0.11)	(0.09)	(0.09)	(0.09)
Is married	0.17	0.12	0.13	0.14
	(0.11)	(0.09)	(0.09)	(0.09)
Bequest (likelihood)	0.00	0.00	-0.00	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Bequest × Frailty (Q1)	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)
Bequest × Frailty (Q2)	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Bequest × Frailty (Q3)	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Bequest × Frailty (Q4)	-0.01	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-2.89***	-2.78***	-2.51***	-2.65***
	(0.75)	(0.56)	(0.55)	(0.55)
Observations	788	1368	1349	1339

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11 Claiming likelihood and health expectations (Logit Specification)

	Any	10k	100k	500k
Frailty (Q2)	-1.45	-2.10	-2.69*	-2.46
	(2.12)	(1.62)	(1.60)	(1.61)
Frailty (Q3)	9.56***	9.13***	9.30***	9.38***
	(1.76)	(1.31)	(1.29)	(1.30)
Frailty (Q4)	1.91	1.22	0.54	0.80
	(2.09)	(1.60)	(1.57)	(1.58)
Frailty (Q1) × Frailty expectations	0.10***	0.10***	0.10***	0.11***
	(0.03)	(0.02)	(0.02)	(0.02)
Frailty (Q2) × Frailty expectations	0.14***	0.16***	0.16***	0.16***
	(0.04)	(0.03)	(0.03)	(0.03)
Frailty (Q3) × Frailty expectations	-0.13***	-0.12***	-0.12***	-0.12***
	(0.03)	(0.02)	(0.02)	(0.02)
Frailty (Q4) × Frailty expectations	0.05*	0.06***	0.06***	0.06***
	(0.03)	(0.02)	(0.02)	(0.02)
Sex	0.13	-0.01	-0.00	0.00
	(0.16)	(0.12)	(0.13)	(0.13)
Liquid Wealth	-0.18*	-0.09***	-0.09***	-0.07*
	(0.10)	(0.04)	(0.04)	(0.03)
Mortality risk	-1.36	-0.11	-0.15	-0.28
	(2.11)	(1.76)	(1.80)	(1.76)
Has college	-0.20	-0.35**	-0.29**	-0.24*
	(0.18)	(0.14)	(0.14)	(0.14)
Is married	0.27	0.19	0.21	0.23
	(0.19)	(0.14)	(0.14)	(0.14)
Bequest (likelihood)	0.01	0.00	-0.00	-0.01***
	(0.01)	(0.00)	(0.00)	(0.00)
Bequest × Frailty (Q1)	0.00	0.00	0.00	0.00
	(.)	(.)	(.)	(.)
Bequest × Frailty (Q2)	-0.01	-0.00	-0.00	0.01
	(0.01)	(0.01)	(0.00)	(0.00)
Bequest × Frailty (Q3)	-0.00	-0.00	-0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.01)
Bequest × Frailty (Q4)	-0.01	-0.01	0.00	0.01
	(0.01)	(0.00)	(0.00)	(0.01)
Constant	-4.64***	-4.51***	-4.10***	-4.37***
	(1.22)	(0.93)	(0.90)	(0.93)
Observations	788	1368	1349	1339

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12 Estimation Results for Probit Regression

	Death indicator
Age	-0.842*
	(0.444)
Age^2	4.182***
	(0.450)
Frailty	2.567***
	(0.124)
$Frailty^2$	-0.352**
	(0.146)
Years of Education	0.002
	(0.002)
Constant	-3.328***
	(0.112)
Observations	205,120

Note: Age is scaled such that: Age = $\frac{(\text{Age}-25)}{100}$.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 13 Estimation of the Frailty Process

	Log of Frailty
Age	0.622
Age^2	2.875
Age^3	-1.032
Age^4	3.556
Constant	-1.866
Persistence	0.944
Variance of Persistent Shock	0.035
Variance of Transitory Shock	0.006
Variance of Fixed Effect	0.737

Note: Age is scaled such that: Age = $\frac{(\text{Age}-25)}{100}$.

Table 14 Estimation of the Income Process

	Log of income
Age	-11.420***
	(0.337)
Age^2	6.675***
	(0.376)
Education	0.369***
	(0.003)
Frailty	-2.611***
	(0.068)
$Frailty^2$	1.455***
	(0.099)
Constant	3.567***
	(0.316)
Observations	199,521
Cohort Effects	Yes
Persistence	0.932
Variance of Persistent Shock	0.233
Variance of Transitory Shock	0.001
Variance of Fixed Effect	0.023

Note: Age is scaled such that: Age = $\frac{(\text{Age}-25)}{100}$.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 15 Estimation of out-of-pocket medical expenses.

	Log of out-of-pocket medical expenses
age	2.859
	(2.869)
age^2	-1.180
	(2.815)
frailty	2.634***
	(0.858)
$frailty^2$	-6.277**
	(2.732)
$frailty^3$	5.155**
	(2.472)
education	0.182***
	(0.012)
Constant	-4.138***
	(1.467)
Observations	7,434
Cohort effects	Yes

Note: age is scaled such that: $age = \frac{(age-25)}{100}$

*** p<0.01, ** p<0.05, * p<0.1

Variance of transitory shock	.073