NBER Retirement Research Center Working Paper
NB04-05

How Fast Should the Social Security Eligibility Age Rise?

David M. Cutler, Jeffrey B. Liebman, and Seamus Smyth
Harvard University and NBER

July 23, 2006

Abstract

Many Social Security reform proposals recommend increasing the age at which people become eligible for Social Security retirement benefits as a way to reduce future expenditures, maintain benefit adequacy in light of other benefit cuts, and increase labor supply. But the amount by which the benefit eligibility age is increased in reform plans is rarely chosen based upon principles of optimal social insurance. In this paper we analyze the question of the optimal eligibility age using two models of optimal social insurance. In the first model, the eligibility age exists because of a paternalistic concern that some people will mistakenly retire too early if left to make decisions on their own. In the second model, the eligibility age represents the age at which it is no longer cost effective to screen people for disability. We parameterize the models using new estimates of how the distribution of health status at each age has evolved in the U.S. over the past four decades.

This research was supported by the U.S. Social Security Administration through grant #10-P-98363-1-01 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER. We thank Eric French for helpful comments on an earlier draft and Mark Shepard for excellent research assistance.

1. Introduction

Many Social Security reform proposals recommend increasing the age at which people become eligible for Social Security retirement benefits as a way to reduce future expenditures, maintain benefit adequacy in light of other benefit cuts, and increase labor supply. For example the Personal Security Accounts (PSA) plan of the 1994-1996 Advisory Council on Social Security would have raised the earliest eligibility age (EEA) from 62 to 65 by 2035 and the 1998 National Commission on Retirement Policy (NCRP) plan would have raised the EEA to 65 by 2017 and then to higher ages thereafter.

The amount by which the benefit eligibility age is increased in reform plans is rarely chosen based upon principles of optimal social insurance. Instead, proposals motivate these changes as holding constant the number of years receiving benefits (the NCRP plan) or holding constant the ratio of retirement years to work years (the PSA plan). These ad hoc targets are unlikely to correspond to what would be recommended by a model of optimal social insurance.[1]

In this paper we analyze the question of the optimal retirement age using two models of optimal social insurance. Both models have at their foundation the decision problem of a representative forward-looking consumer and, in particular, the comparative statics of how this consumer's choice of retirement date is affected by rising productivity, a longer lifespan, and lower disability rates at each age. Rising productivity will lead to retirement at an earlier age if the income effect of higher wage levels outweighs the substitution effect. Longer lifespans will increase the marginal utility of income, leading

[1] Two scholarly reform proposals do apply logic similar to that of this paper in discussing reform options. Aaron and Reischauer (2001) recommend raising the EEA, but note that the policy will create winners and losers. Their proposal would keep the replacement rate constant for workers who spend the same fraction of their adult lives in retirement as those reaching age 62 in 2011. Diamond and Orszag (2004) do not recommend raising the EEA, arguing that population heterogeneity makes it difficult to judge whether the benefits of doing so would outweigh the costs.

to later retirement. Lower disability rates will reduce the disutility of work and also lead to later retirement. The two models then extend this representative consumer analysis by introducing heterogeneity (in wages, retirement benefits, and initial period assets).

Importantly, one needs a rationale for Social Security to answer the eligibility age question -- why people are not simply permitted to tap their Social Security benefits at whatever age they choose with an appropriate actuarial adjustment for early receipt of benefits. We consider two rationales. In the first model, the eligibility age exists because of a paternalistic concern that some people will mistakenly retire too early if left to make decisions on their own. The choice of an initial age at which people can access retirement benefits therefore involves balancing a desire to force myopic people to keep working against the cost of not permitting people for whom it would be optimal to receive benefits before the retirement age to do so. In the second model, the eligibility age is set by considering the interaction between Social Security and Disability. If all we wanted to insure were health status, we would have a disability program but no Social Security. In practice, screens for disability are imperfect and costly. At some age the costs of screening outweigh the benefits. This is the age at which universal access to Social Security should begin.

We parameterize our models using new estimates of how the distribution of health status at each age has evolved in the U.S. over the past four decades. These estimates allow us to make statements about how fast the eligibility age should rise in the future if health gains in coming decades are similar to what we experienced in the past. More generally, our health estimates allow us to make statements such as "70 year old males

today are as healthy as 62 year olds were in 1961." This information is of use in thinking about eligibility-age policy even in the absence of our specific modeling assumptions.

2. Background on the Social Security Eligibility Age

The Social Security Act of 1935 set age 65 as the age at which people could first receive benefits. The 1889 German contributory pension system and the 1908 British means-tested system had both limited eligibility to people age 70 and above (Rodgers 1998).[2] In the U.S., however, the 1890 pension system for Union army veterans and many other private and public pension plans established in the early 20th century set age 65 as the age for benefit entitlement (Costa, 1998). Indeed, Wilbur Cohen a staff member of Roosevelt's Committee on Economic Security has written that "at no time in 1934 or 1935 did President Roosevelt, the Committee on Economic Security, or the key members of Congress handling the bill deem feasible or appropriate any age other than 65 as the earliest eligible age for receipt of old-age 'insurance' benefits."[3]

The age at which people could first receive retirement benefits, termed the Early Eligibility Age or EEA, was reduced to 62 for female retirees beginning in 1956. Proponents of this change argued that women were on average several years younger than their husbands and deserved to be able to retire at the same time (Meriam and Schlotterbeck, 1950, p. 22). Older women were also seen as facing difficulties in the labor market. Then, in February 1961, President Kennedy proposed Social Security benefit expansions as part of his initial economic package. These proposals were designed to address the weak employment situation and included a recommendation to

[2] Germany reduced its retirement age to 65 in 1916. However, this does not seem to have influenced the architects of the U.S. system. See http://www.ssa.gov/history/age65.html.
[3] Cohen (1985), p. 10.

lower the EEA for men to the same level as for women. The EEA reduction for men was approved by Congress and went into effect in late 1961.[4]

Over time, a greater fraction of workers have taken advantage of the opportunity to claim benefits before age 65. In 1965, a few years after men first became eligible to claim at age 62, 16.6 percent of men claimed benefits at age 62, 17.0 percent claimed benefits at ages 63 or 64, and 66.5 percent claimed benefits at ages 65 or older. By 1975, 29.2 percent claimed benefits at 62, 26.1 percent at ages 63 or 64, and 44.7 percent at ages 65 or older. Today, 54 percent claim benefits at age 62, 22 percent at ages 63 or 64, and only 25 percent at ages 65 or above.[5]

There are several other Social Security policy parameters related to retirement, including the full retirement age -- the age at which one's benefits are not reduced for early retirement -- and the actuarial adjustment associated with retirement at different ages.[6] Between 2000 and 2022, the full retirement age is to rise from 65 to 67. Workers who choose to receive benefits before the full retirement age have their benefits reduced according to the actuarial adjustment factor of 6.7 percent per year.[7] Workers who claim benefits after the full retirement age receive a delayed retirement credit that increases their benefits by 8 percent per year. While historically there was a strong financial incentive to retire by age 65, for a worker reaching age 62 today, these parameters are

[4] We note that by repeatedly alternating these two rationales for adjusting the retirement age, we could quickly converge on an EEA of zero.

[5] Diamond and Orszag (2004). These numbers exclude people converting from disability benefits to retirement benefits at the full benefit age. There have been similar trends for females.

[6] Though the age of full benefit entitlement is sometimes referred to as the normal retirement age (NRA), Henry Aaron has pointed out that the term "normal retirement age" does not appear anywhere in the Social Security Act or regulations. Moreover, there is nothing "normal" about it since two-thirds of people claim benefits before that age. Thus we join his crusade to eliminate this misleading term from discussions of Social Security policy. See Diamond and Mirlees (1978) and Diamond and Mirlees (1986) for models of how benefits should rise with a worker's retirement date.

[7] There is a slightly smaller adjustment for people who retire more than 36 months before the full benefit age.

approximately actuarially fair on average.[8] Other provisions including the payroll tax rate, benefit calculation formulae, and earnings test affect retirement incentives as well.[9]

It is worth emphasizing that an increase in the full retirement age represents a reduction in aggregate benefit payments. In contrast, raising the earliest eligibility age has approximately no effect on system finances since the benefits received later are increased to adjust for their delayed receipt by an amount that is roughly actuarially fair.[10] While a full assessment of retirement-age policy would analyze tax and benefit levels and the rate at which benefits rise with age when claiming is delayed, our emphasis in this paper is on the age at which benefits availability begins, and therefore our policy experiments correspond to changing the EEA while holding all other parameters of the Social Security system constant. For that reason, the changes we study keep expected lifetime benefit levels constant.[11]

3. Theory

To analyze optimal eligibility age policy for Social Security, we need to combine a model of retirement decisions by individuals with a rationale for government provision of retirement income. In this section we first describe the model of individual decision making and show some comparative statics for how an individual's choice of a retirement

[8] Coile and Gruber (2001) show, however, that there is substantial heterogeneity in incentives across individuals.

[9] For people ages 65 and above, there is no necessary connection between benefit claiming and retirement. For people between 62 and 64 who continue to work, benefits are reduced via an earnings test (though the lost benefits are repaid in the form of higher later benefits). On the distinction between benefit claiming and retirement see Coile, Diamond, Gruber, and Jousten (2002).

[10] In this discussion, we ignore impacts of changes in the early eligibility age on income tax revenues.

[11] This statement is only approximately true. Our analysis holds expected lifetime benefits constant for each individual. An actual change in the EEA would hold expected lifetime benefits constant for a cohort overall, but if an individual's mortality risk differed from the cohort average, there would be distributional and incentive effects of a change in the EEA that are not captured in our model.

age varies with productivity growth, increases in longevity, and improvements in health.[12] Then we combine this model with two different rationales for Social Security – one that hinges on worker myopia and another that hinges on the costs of verifying disability status.

The Individual's Retirement Decision

We model the retirement decision as a binary, annual, irreversible, choice. At each age, t, every person who is still working faces a decision of whether or not to keep working in that year as well as a decision about how much to consume. If the individual works he or she receives additional income, w_t. The retirement and consumption decisions are made to maximize expected utility over the worker's remaining lifetime:

$$V_t = \max_{l_t, c_t} E\left[\sum_{t=current\ age}^{A_{\max}} \frac{[v(c_t) - e(h_t, t)l_t]p_t}{(1+r)^{t-current\ age}} \right] \quad (1)$$

Utility in each period is an additively separable function of the utility of consumption, $v(c_t)$, and the disutility of labor effort, $e(h_t,t)l_t$.[13] The disutility of labor effort rises as health status declines and is zero if a person does not work ($l_t=0$). It may also depend on age for reasons other than health. p_t is the probability that an individual survives to age t. r is the discount rate. We assume that the only sources of uncertainty are health status and age of death. In particular, the (age-varying) wage that will be received by the worker in each year if the worker continues to work is known in advance, as are

[12] Bloom, Canning, and Moore (2004) analyze similar issues in a model that lacks a rationale for Social Security. See also Sheshinski (1999).

[13] Although we follow conventional practice in assuming that utility for each year is additively separable, this assumption is not innocuous. For example, there may be reasons other than declining health for why people bunch their leisure at the end of their lifetime. Some uses of leisure (traveling around the world, visiting grandchildren who live in another city) may require long contiguous blocks of leisure. In this case, the marginal value of leisure could depend on total lifetime leisure rather than simply on leisure in the current period.

retirement benefits (which are assumed to vary with the date of retirement in a way that is known in advance).

The expected future utility at time t can be expressed as a function of the three state variables: beginning of period assets, an indicator for whether the person enters the period retired, and an indicator for whether the person is in good health:

$$V_t = V\left[a_t, I(retired_t), I(health_t)\right] \qquad (2)$$

An individual's choice of retirement date and the comparative statics associated with changes in the economic environment can usefully be illustrated by considering an individual's expected retirement age viewed from the first period of the model (in our empirical work this will be age 55).

For each potential retirement age, R, it is possible to calculate the difference, $D_{R,}$ between the expected value of continuing to work through that year (and possibly continuing to work in subsequent years) and the expected value of working through the prior year, but retiring in year R. An individual's expected retirement date, R^*, will be the age at which this difference first becomes negative.

This difference can be decomposed into three components. First, by working an extra year a person obtains extra income (including the after-tax wage for that year and additional entitlement to pension and Social Security benefits). The extra income results in higher consumption and therefore higher utility in the current year and in all future years. This component of the difference can be calculated as the difference in value between coming into the period with the initial asset level that would occur if the person planned to retire in that period and coming in with additional assets equal to the incremental income obtained by working in that year.

$$EV_R\left[a_R + \frac{w_R}{1+r}, 1, I(health_R)\right] - EV_R\left[a_R, 1, I(health_R)\right] \qquad (3)$$

Second, by continuing to work a person maintains the option of working in future. The utility benefit of the option value is:

$$\frac{1}{1+\rho}\frac{p_{R+1}}{p_R}\left\{EV_{R+1}\left[a_{R+1}, 0, I(health_{R+1})\right] - EV_{R=1}\left[a_{R+1}, 1, I(health_{R+1})\right]\right\} \quad (4)$$

Third, the person incurs the disutility of effort for the current year's labor, $E[e(h_R,R)]$. If the expected consumption benefits and option value outweigh the expected disutility of effort, then the individual will expect to work rather than retire in year R.

Figure 1 illustrates the three components of the individual's expected retirement decision, using a simulation from the dynamic programming implementation of this model that is described below. Curve AA shows the value of the additional year of income. As retirement is postponed to later years, the value of an additional year of income falls because lifetime income is higher and therefore the marginal utility of additional consumption is lower.[14] Line CC adds the option value of continuing work to the value of additional income – and thus represents the total benefits that accrue from working an additional year. Option value is greatest at earlier ages and falls to zero just before the age where the person is certain to retire regardless of health status. Line BB represents the disutility of from an additional year of effort which is assumed to rise with age as health deteriorates. The individual expects to retire at age R^* -- where the CA and BB curves intersect.

[14] At older ages w_t will also likely fall with age.

Consider now how the curves will shift over time. We simplify by ignoring option value – which is small near the optimal retirement age.[15] Increases in productivity raise wage levels, which results in offsetting income and substitution effects. Thus it is theoretically ambiguous whether the marginal utility of income curve will shift left or right. The shift from line AA to line AA' in Figure 2 shows an example in which the income effect dominates and so desired retirement ages will fall.

As longevity increases (holding health at each age and the wage level constant), consumption per year of life will fall as there will be more years over which to spread one's income. This change will shift the marginal utility of income curve to the right and result in later retirement. Note, however, that the net impact of the income effect and the longevity effect is likely to be a shift in the AA curve to the left. Life expectancy at age 20 has increased by only about 0.25 percent per year since 1960, while real wage growth has been roughly 1 percent per year. Thus, on average, resources have expanded more rapidly than the number of years they need to be spread over. Nonetheless, if the substitution effect of higher wages is sufficient, the AA curve could shift to the right.

As health at each age improves over time due to advances in medical care and changes in health behaviors, the disutility of effort falls, and the BB curve will shift downward and to the right (to BB' in figure 2).[16] The net effect of income and health changes is thus likely to be a shift to the left of the marginal utility of income and a shift

[15] In the model this is true because negative health shocks at older ages tend to persist and because health status tends to shift the retirement date by only a few years. In the real world there is an additional reason for option value to be small – retirement decisions are not completely irreversible.

[16] Changes in job requirements over time could also shift the disutility of effort curve. If jobs require less manual effort, the curve will shift further to the right. On the other hand, workplace changes such as increased cognitive demands or the increase in two-worker households could increase the disutility of work effort over time.

to the right of the marginal disutility of effort. Both the magnitude of the shift in the individual's optimal retirement date and also the direction is theoretically ambiguous.

The analysis so far has described the behavior of an individual worker. More generally, there will be a separate set of curves for each worker, indexed by wage level, initial-period assets, and health, and our analysis of retirement policy will involve summing social welfare over the full distribution of worker types. Before doing this, however, we need to discuss rationales for Social Security.

Social Security Eligibility Age as a Cure for Myopia

In our first model, the rationale for the Social Security eligibility age is to prevent myopic people from accessing their benefits too early and thereby ending up with too low a consumption level in their later years. There are several different channels through which the myopia might operate. Myopia could operate in the standard way through individuals discounting the future at too high a rate (perhaps hyperbolically). In that case, myopic individuals would tend to be the same individuals who have inadequate private saving. Alternatively, otherwise forward-looking individuals might simply retire at the first date Social Security benefits are available or the first date private pension benefits are available rather than going through the effort of optimizing, perhaps assuming that the government or their employer has done the optimization for them. In this case, there might be little correlation between excessively early retirement and savings levels.

Let R^*_i be the optimal retirement age for each individual if the individual were forward-looking and could gain access to their retirement benefits at any age. Let R_{age} be

the government-set age at which people can first access their benefits (with actuarially fair adjustments for delays in claiming). If credits for delayed retirement are actuarially fair,[17] forward-looking people with sufficient assets will not be affected by R_{age}; they will borrow and lend to smooth income around the age at which they can collect benefits. Two other types of individuals will be affected by Rage. First, some forward-looking individuals with low assets may want to access their retirement benefits at a relatively young age. These people will be forced to work longer than is optimal, and their well-being will decline as the eligibility age rises. Second, some myopic individuals would optimally work longer than the earliest eligibility age, and these people give up additional income by retiring earlier. As the eligibility age rises to a level closer to a myopic person's optimal retirement age, the well-being of these individuals improves. The optimal eligibility age balances these two considerations, weighted by the number of people and average per-person utility loss of retiring at a suboptimal age in each case.

Social Security as Insurance Against Adverse Health Shocks

Our second model interprets the Social Security retirement age as the age at which workers no longer need to meet a disability test to receive benefits. To analyze the issue from this perspective, we extend a model of Diamond and Sheshenski (1995) to incorporate costs of screening for disability.[18]

To begin, consider a population of workers with identical preferences and a distribution of wages, *f(w),* that ranges from $\underline{w}$ to $\overline{w}$. In each period a worker has the

[17] I.e. the expected present value of retirement benefits is independent of the age at which one claims them.
[18] Waidman, Bound, and Nichols (2003) also extend Diamond and Sheshinski (1995) to incorporate a more realistic model of the disability screening process. Their focus, however, is on variation in the screening mechanism and they do not address the issue of choosing an optimal retirement age.

option of applying for disability insurance. The government screens workers who apply for disability. This screen costs the government A and the individual undergoing the screening B. Screening is imperfect. Thus agents with poor health are successful at claiming disability with probability $p_p>1/2$, while those in good health succeed with probability $p_g<1/2$. The probability of being in poor health at each age, t, is ϕ_t which rises with age.

Workers receive utility from consumption and leisure. $u(c_w)$ is consumption from working and $u(c_d)$ is consumption from receiving disability insurance. While working, agents incur disutility from effort that varies with their health status: $v(h_p)>v(h_g)$.

An agent in poor health will apply for disability insurance if

$$(1-p_p)\left[\left(u(c_w)-B\right)-v(h_p)\right]+p_p\left[u(c_d)-B\right]>u(c_w)-v(h_p)\,. \qquad (5)$$

This simplifies to:

$$p_p\left[u(c_d)-(u(c_w)-v(h_p))\right]>B\,. \qquad (6)$$

The analogous condition for a healthy agent is:

$$p_g\left[u(c_d)-(u(c_w)-v(h_g))\right]>B \qquad (7)$$

Define $\overline{w}_p$ and $\overline{w}_g$ as the wage levels at which expressions (6) and (7) hold with equality for agents in poor health and good health respectively. I.e. at wages above these levels people do not apply for disability, whereas at wages below these levels people do apply. Because $v(h_p)>v(h_g)$, we know that $\overline{w}_p > \overline{w}_g$.

If the government does not screen for disability, workers in poor health will apply for benefits whenever:

$$u(c_d)-\left[u(c_w)-v(h_p)\right]>0 \qquad (8)$$

and workers in good health will apply for benefits whenever

$$u(c_d) - \left[u(c_w) - v(h_g)\right] > 0 \,. \tag{9}$$

Define $\overline{w}_{np}$ and $\overline{w}_{ng}$ as the wage levels at which expressions (8) and (9) hold with equality -- at wages above these levels people do not apply for benefits, even when the government does not screen for disability. Assume that income transfers create deadweight loss and therefore the social cost of raising a dollar of revenue is MCF>1.

Total social welfare with screening is

$$\phi_t\left[\int_{\underline{w}}^{\overline{w}_p}\left\{p_p[u(c_d) - MCFc_d] + (1-p_p)[u(c_w) - v(h_p)] - B - A\right\}f(w)dw + \int_{\overline{w}_p}^{\overline{w}}\left\{u(c_w) - v(h_p)\right\}f(w)dw\right] + \tag{10}$$
$$(1-\phi_t)\left[\int_{\underline{w}}^{\overline{w}_g}\left\{p_g[u(c_d) - MCFc_d] + (1-p_g)[u(c_w) - v(h_g)] - B - A\right\}f(w)dw + \int_{\overline{w}_g}^{\overline{w}}\left\{u(c_w) - v(h_g)\right\}f(w)dw\right]$$

and total social welfare without screening is:

$$\phi_t\left[\int_{\underline{w}}^{\overline{w}_{np}}\left\{[u(c_d) - MCFc_d]\right\}f(w)dw + \int_{\overline{w}_{np}}^{\overline{w}}\left\{u(c_w) - v(h_p)\right\}f(w)dw\right] + \tag{11}$$
$$(1-\phi_t)\left[\int_{\underline{w}}^{\overline{w}_{ng}}\left\{[u(c_d) - MCFc_d]\right\}f(w)dw + \int_{\overline{w}_{ng}}^{\overline{w}}\left\{u(c_w) - v(h_g)\right\}f(w)dw\right]$$

The difference in social welfare between screening and not screening consists of three components. First, under screening the government and applicants bear the cost of screening:

$$\textbf{Screening costs} = \phi_t\int_{\underline{w}}^{\overline{w}_p}(B+A)f(w)dw + (1-\phi_t)\int_{\underline{w}}^{\overline{w}_g}(B+A)f(w)dw \tag{12}$$

Second, under screening, some sick individuals who apply for benefits are denied benefits, and some sick individuals are deterred from applying for benefits due to the application cost. Assuming that the utility gain for individuals in poor health who apply

for and receive disability benefits exceeds the marginal cost of funds, then this is a further cost of screening.

Cost of denying benefits to sick=

$$\phi_t\left[(1-p_p)\int_{\underline{w}}^{\overline{w}_p}\{u(c_d)-[u(c_w)-v(h_p)]-MCFc_d\}f(w)dw+\int_{\overline{w}_p}^{\overline{w}_{np}}\{u(c_d)-[u(c_w)-v(h_p)]-MCFc_d\}f(w)dw\right] \quad (13)$$

The benefit of screening is that it prevents healthy individuals from claiming and (with probability p_g) receiving benefits. Assuming that on average the utility gain for individuals in good health who apply for and receive disability benefits is less that the marginal cost of funds, then this benefit is:[19]

Benefit from denying payments to healthy=

$$(1-\phi_t)\left[(1-p_g)\int_{\underline{w}}^{\overline{w}_g}\{MCFc_d-u(c_d)+[u(c_w)-v(h_g)]\}f(w)dw+\int_{\overline{w}_g}^{\overline{w}_{ng}}\{MCFc_d-u(c_d)+[u(c_w)-v(h_g)]\}f(w)dw\right] \quad (14)$$

At older ages, the fraction of the population in poor health, ϕ_t, rises. This will raise screening costs since it raises the fraction of the population that apply for benefits and need to be screened (recall that $\overline{w}_p > \overline{w}_g$). It will also raise the number of sick people who are denied benefits. So both costs of screening rise with age. Moreover, the benefits of screening – denying payments to the healthy – decline with age as there are fewer healthy people who can potentially claim benefits. Thus at some age the costs of screening will exceed the benefits of screening and at that age universal entitlement to retirement benefits should begin. As the population's health improves over time, ϕ_t at each age will fall. This will make screening worthwhile up to higher ages than before – raising the optimal age of universal entitlement to retirement benefits.

[19] It is possible that some healthy people have low enough wage levels so that giving them disability benefits represents an increase in social welfare.

4. Changes in the distribution of health

In this section, we present trends in the health of the elderly and near-elderly population, to identify how health has changed over time. Health is, of course, multi-dimensional. While there are indices that try to combine various measures of health typically on a 0 to 1 quality of life scale, we consider instead basic descriptive data on several important dimensions.

One can think about health differences at the high end of functioning or at the low end. At the high end, there are questions about just how vigorously a person can exercise, and what other strenuous activities they can engage in. At the low end, questions revolve around whether the person can care for themselves, undertake the normal activities of life, and function independently. For understanding the ability to work, the relevant issue is health at the bottom end of the distribution. We focus on that dimension.

Mortality is the end point of poor health, so it is natural to begin there. Figure 3 shows the share of men at each age who are near death – which we define as within two years of dying.[20] Two years prior to death is a period in which disability rates are very high. We report results separately for men in 1960 and men in 2000. Because of our interest in social security retirement ages, we focus on the population of men in their early 60s.

In 1960, the two-year mortality rate for men in their low 60s was about 6 percent. In 2000, the mortality rate was over 40 percent lower. For our purposes, we care less

[20] These data are from US life tables and represent a portrait of what the typical man of each age would experience in that year.

about changes in mortality at each age than in what we term disability-adjusted age. We ask the question: at what chronological age today would one be in the same health as a person of a given age 40 years previously? An increase in disability-adjusted age is equivalent to an improvement in health for that age group. Figure 3 shows that the average 68 year-old in 2000 has the same probability of death in the next two years of life as the average 62 year-old in 1960. Thus, disability-adjusted age has increased by 6 years in the past four decades.

A similar analysis is true about life expectancy. The life expectancy of a 62 year-old in 1960 was 14.6 years. That is the life expectancy today of a person aged 67. Thus, if one were defining a Social Security retirement age based only on expected years until death, the retirement age would have increased by about 5 years in the past four decades.

Mortality is not the only endpoint, of course; non-fatal health may be even more important. We consider a number of dimensions of non-fatal health. The first is self-reported general health status. Since the early 1970s, The National Health Interview Survey has asked a question of the form: "Would you say your health in general is excellent, very good, good, fair, or poor?"[21] The typical measure of poor health is whether the person answers fair or poor. Poor health measured this way is strongly associated with subsequent mortality (Idler and Kasl, 1991). To ensure adequate sample size, we aggregate NHIS responses for two time-periods: 1974-76 and 1994-96. The former years are the earliest years of data availability; the latter years are towards the end of the data that have been released.

Figure 4 reports the share of men in fair or poor health over time. The jagged line is the raw data; the trend line is formed as the predictions from a quadratic in age. In the

[21] The exact wording of the question has changed over time, but the meaning has been virtually the same.

mid-1970s, 29 percent of men aged 62 reported their health as being fair or poor. Over time, health has improved, especially at older ages. By the mid-1990s, one did not reach the same share of men reporting fair or poor health until ages in the low 70s. By this measure, health has improved by about 10 years.[22]

It is notable that even with the fewer number of years in the intervening period, the gains in health status are larger than the gains in longevity. Healthy life is rising more rapidly than are years of life.[23] This is particularly impressive because one would expect some tendency for reported health to remain stable over time if people's expectations about what they should be able to do increase with improvements in physical functioning.

In addition to self-reports, one would like to compare direct physical assessments of health – for example, whether the person can stand for a certain length of time or perform essential cognitive tasks. Some surveys of health do perform physical assessments, but the questions asked are not the same over time. An alternative is to look at individual reports of impairment in specific domains. The National Health Interview Survey asks people who report having a variety of different conditions how many days the condition caused the person to be in bed for all or most of the day.[24] We tabulate responses for people who report having a stroke, heart attack, ischemic heart disease, or other related conditions (ICD-9 codes 410-414 and 430-438). These conditions are common and relatively severe. In this case, severity is a virtue: numerous studies show that doctors diagnose less severe conditions more than they used to, and hence the set of

[22] Another NHIS question asks whether people are limited in the kind or amount of their major activity. The difficulty with this question is that peoples' major activity varies over time. The share of people limited in their major activity, for example, declines with increasing age after retirement.

[23] This is consistent with the observation that disability rates among the young elderly are falling (Cutler, 2001).

[24] In 1972, the specific question is about how many days the person was in bed for all or most of the day. In 1994-96, the question is about how many days the person was in bed at least half the day.

people with less severe conditions such as arthritis or back pain would not be the same over time.[25] Heart disease diagnoses, in contrast, seem to have matched trends in clinical measurement of disease incidence well (Cutler and Richardson, 1997).

Figure 5 shows annual bed days for people with these conditions. We start at age 45, since heart disease is less common below that age. Because our sample sizes are smaller, each individual observation is a three-year moving average. For the same reason, we pool men and women.

Impairment associated with heart disease has fallen over time. In 1972, the average 62 year-old male with heart disease reported 20 days in bed in the previous year. By the mid-1990s, age-specific bed days were about 5 days lower across the age spectrum.

More general measures of disability suggest the same finding. The National Long-Term Care Survey has conducted assessments of disability among the elderly over the course of the past two decades – defined as the ability to perform basic activities of independent life such as dressing, bathing, and doing light housework. The share of the 65 to 74 year-old population that is disabled in this way fell from 12 to 9 percent.

Considering all the evidence, it is clear that health near traditional retirement ages has improved markedly over time. Our best guess is that people aged 62 in the 1960s or 1970s are in equivalent health to people aged 70 or more today.

[25] To provide a sense of the magnitude of the increased diagnosis of chronic conditions, the average 62 year-old in the mid-1970s reported about 0.7 chronic diseases. In the mid-1990s, that share of diseases was reported by the average 47 year-old.

5. Empirical Implementation of the Myopia Model

Dynamic Programming Approach

We implement a dynamic programming model using data from the Health and Retirement Survey to study the implications of the theory discussed in section 3 in a more realistic setting. We build upon two recent papers that estimate similar models and explore the behavioral responses to changing the Social Security eligibility age. Our approach is most similar to Gustman and Steinmeier (2005). In particular, we use the same specification for preferences, the same simulated method-of-moments estimatation technique, and the same HRS data set. Unlike Gustman and Steinmeier, who treat only one's mortality date as uncertain, we treat both future mortality and health as uncertain; therefore our dynamic programming problem is more complicated than theirs. As a tradeoff, we make one major simplification relative to Gustman and Steinmeier in that we do not permit a part-time work option. French (2005) estimates a life-cycle model of labor supply, retirement, and savings behavior in which future wages, health, and mortality are uncertain. French's approach is not suitable for our purposes because it does not allow for heterogeneity in preferences or heterogeneity in the data generating process for the state variables. Moreover, French uses non-separable preferences that constrain the income and substitution effect of an increase in wage levels to cancel. Such a specification of preferences rules out the possibility that the AA curve in figure 1 shifts to the left, something we would like to be able to explore.

In particular we solve the following maximization problem for each individual:

$$\underset{l_t c_t}{Max}\left[\frac{c_t^{1-\gamma}}{1-\gamma}+(1-l_t)e^{\beta_0+\beta_{age}(age-62)+\beta_{health}(bad\ health)+\varepsilon_i}\right]+\beta\phi E\left[V(h_{t+1}.a_{t+1},I(retired)\right]$$

$$h_{t+1}=P(|\,h_t)$$

$$a_{t+1}=(1+r)a_t+w_t l_t-c_t$$

$$a_{t+1}\geq 0$$

where l_t is an indicator that equals one if the person works in year t. In words, each individual chooses consumption and decides whether to work or retire in the current period so as to maximizes the utility of the current period and discounted expected future utility. Current period utility is the sum of a consumption term and a term that gives the utility of leisure. The latter is a function of health status, since the disutility of work is greater with worse health. The future is discounted for two reasons: time preference, β, and the chance of mortality, ϕ. There are three state variables in the model: health status, financial assets, and whether the individual has previously made the irreversible decision to retire. Health status evolves according to a Markov process. Asset levels are determined by a standard asset accumulation equation and are constrained to be non-negative.

We solve the model from age 55 to age 100 using backward induction. We assume that everyone alive at 100 automatically dies at that age. Health status and retirement status are both discrete. In calculating the value function, we discretize over possible asset levels and consumption choices, interpolate using Akima's shape preserving spline, and find the maximum using a non-linear interior point solver maximization algorithm contained in the OPT++ toolkit. In order to let wages vary with age and pension benefits to vary with age of retirement, we solve the model separately for each individual in our data set.

Data

Our data come from the RAND Health and Retirement Survey (HRS) version D and the HRS Imputed Data for Pension Wealth Calculations Version 1.0 for males in the original HRS cohort – born between 1931 and 1941.

The moments for our simulated methods of moments estimation procedure are the fraction of sample members at each year of age from 54 to 68 who are retired and in good health, retired in bad health, working in good health, and working and in bad health. Good health is defined as health that is excellent, very good, or good. Bad health is health that is fair or poor. A person is retired if their self-reported retirement status is fully retired or if self-reported retirement status is partially retired and earnings are below $10,000. To estimate the true moments, we use the entire data set. The number of observations per year of age ranges from 569 (at age 68) to 2256 (at age 59). To keep computation costs reasonable (i.e. to allow the model to be estimated in 8 hours using four processors operating in parallel), we calculate our simulated moments using a random 132 person subset of the data.

Each worker's annual wage is taken from the wage in the 1992 HRS survey and is assumed to remain constant. Social Security benefits are calculated by applying the benefit formula to an imputed level of AIME based on 1992 earnings.[26] Defined Benefit pension benefits are taken from the pension benefit file described above and benefit levels are interpolated for years between years available in that data set. Defined contribution pensions and other non-housing financial assets are assumed to earn a 2.3

[26] We have recently applied for access to the restricted version of the HRS with SSA earnings records so as to improve the accuracy of our Social Security benefit projections.

percent rate of return and can be consumed at any time (Social Security benefits and pension benefits cannot be consumed until retirement). Mortality data come from Social Security Administration cohort mortality projections for the 1940 and 1900 birth cohorts. The probability of transitioning between good and bad health states is estimated from two-year health transitions in the HRS.[27] Following Samwick (1998) and Gustman and Steinmeier (2005), we assume that 40 percent of the population has a discount rate of 2.5 percent, 21 percent has a discount rate of 7.5 percent, 6 percent has a discount rate of 12.5 percent, and one-third has a discount rate of 80 percent. We have estimated our model with these discount rates correlated with the initial asset-to-wage ratio and with it uncorrelated, and the two methods produced very similar estimates.[28]

Parameterizing and Estimating the Model

The model has five parameters: γ, β_0, β_{age}, β_{health}, and σ_{ε}.. Using only the moments described above, some pairs of parameters are difficult to separately identify. In particular, as can been seen by referring to figure 1, γ and β_0 determine the location of the AA and BB curves and therefore jointly determine the average retirement age. However, many pairs of values of these two parameters generate essentially the same moments. Thus, some additional information needs to be brought in to identify these parameters. Gustman and Steinmeier (2005) identify γ=1.26 using the differential retirement rates of high and low wage individuals. This is an appealing approach since it identifies the parameters from the same data as the rest of the estimation. There is of

[27] We thank Amy Finkelstein for calculating these transition probability for us.
[28] Once we obtain access to the restricted HRS we will follow a procedure like that of Gustman and Steinmeier to generate individual specific discount rates consistent with lifetime earnings patterns and age 55 wealth.

course a large literature on values for the coefficient of relative risk aversion. Chetty (2003) argues that most microeconomic analysis suggests that γ must be between 1 (log utility) and 2, with his preferred value near 1. With log utility, the income and substitution effects cancel and the AA curve does not shift due to rising wage levels. Since the AA curve will shift right due to rising longevity and the BB curve will shift right from improved health, with log utility we know unambiguously that successive cohorts of forward-looking individuals will choose later retirement dates (all else equal). With γ=2, the income effect will dominate the substitution effect. Simple simulations suggest that with productivity growth of 1 percent per year, this will shift the retirement date down by about 2 to 3 years every four decades (the exact number depends on the slope of the BB curve). When combined with the longevity shift, shifts over time in the AA curve with this parameter will tend to reduce a forward looking individual's retirement age by about 1 to 2 years. Given that the BB curve will shift right, the net effect is theoretically ambiguous. To show the sensitivity of policy conclusions to the value of γ, we fix γ at two values (1.26 and 2) and estimate the other four parameters.

A second identification issue is how to separately identify the effect of health status from other reasons that the disutility of effort might rise with age – given that population health deteriorates smoothly with age. In particular, our binary measure of health status is unlikely to fully capture all health changes. Therefore, the age term in the model is likely to also be capturing some effects of health. In practice, we find the ratio of the health coefficient to the age coefficient to be highly unstable – ranging from 1.0 to over 20 depending on the other parameter values (most of the estimates are between 5 and 10). In our policy simulations, we present both a lower bound estimate of the impact

of health, allowing improvements in health to shift behavior only through the health term (i.e. assuming that the age term has no health-related component) and an upper bound estimate of the impact of health, allowing improvements in health to shift behavior through both the health term and the age term (i.e. assuming that the entire age term is attributable only to health).

A third identification issue is distinguishing among different sources of heterogeneity in retirement dates. With γ and β_0 pinning down the average retirement date, there are four remaining factors that can allow the model to match the empirical spread in retirement dates around the mean as well as the spikes in retirement at age 62 (and to a limited extent) age 65: heterogeneity in tastes (σ_e), the magnitude of the age coefficient (at high values everyone retires at the same date), heterogeneity in discount rates, and heterogeneity in pension plan incentives. We find that a model with only a couple of these sources of heterogeneity can fit the basic retirement/health moments we use almost as well as a model with all four sources of heterogeneity and that the values of the age and heterogeneity parameters vary significantly in models with different amounts of discount rate and pension plan heterogeneity. Thus, further progress in identifying the model will require, at the very least, the use of more moments – for example, matching separate retirement and health status means for those with and without pensions – or incorporating evidence from outside the model such as from natural experiment studies.

Table 1 shows estimates from our model. These estimates are preliminary as we are currently using relatively crude measures of pension and Social Security benefits and have estimated each model with only a few starting values. The table also shows the estimates from Gustman and Steinmeier (2005) for comparison purposes. In our

estimates we are finding a somewhat higher age coefficient than did Gustman and Steinmeier. Our health coefficients are much larger. With γ=1.26, poor health is the equivalent of 12 years of age compared with 4 years in Gustman and Steinmeier. With γ=2.00 our health coefficient is implausibly large. Because relatively few people transition from good to bad health in their mid 60s, are policy simulations are not very sensitive to the estimate of the health parameter. Figure 6 shows the true data moments and the simulated moments for each of our two models.

6. Simulations of Optimal Policies

This section reports results of preliminary policy simulations from each of our two sets of estimates. Figure 7 displays the results of policy simulations using our estimates with γ=1.26. In order to avoid complications associated with conducting welfare analysis when different individuals have different preferences, we conduct these simulations with a discount rate of 2.3 percent for all sample members. We start by solving the model for a sample of forward-looking individuals. Then we augment the sample to include a set of myopic individuals who take the earliest-eligibility age as a cue and retire at the EEA rather than optimizing.

The top left panel of figure 7 shows how average lifetime utility varies with the eligibility age for three different types of individuals. The first are forward-looking individuals with asset to wage ratios (at age 55) of at least 5. We call these individuals "unconstrained" individuals because they have sufficient assets to consume out of if they retire before the EEA so that they are largely indifferent to the EEA. Thus their lifetime

utility is largely unaffected by varying the EEA. The second group of individuals are forward-looking individuals with asset to wage ratios below 5. Some of these individuals would optimally want to retire before the EEA (especially if they are in poor health) if they could tap their Social Security benefits, but are harmed by not being able to do so. For this population it would be best if there were no EEA at all and people could claim benefits (with appropriate actuarial adjustments in their level) at any age they choose.

The third group are the myopes. At low values of the EEA, many of them retire too early and have lower utility than if they were constrained to retire later. As the EEA rises, average utility for this group rises until it reaches a point near the average retirement age that would be desired by myopes if they were forward-looking. At ages above this level, average utility falls as utility losses for people who are being constrained to retire after their optimal retirement age exceed the utility gains for people with high desired retirement ages who are being brought closer to their optimum age.

The optimal eligibility age for the government to set will depend on the population share of these three groups. Since unconstrained individuals are essentially unaffected by the EEA, the key question is the relative size of the myope and constrained forward-looking populations. Three quarters of HRS sample members have asset-to-wage levels at age 55 that are below 5. The bottom left panel of figure 7 shows overall average utility assuming that 50 percent of the population are myopes, 37.5 percent are constrained forward-looking individuals, and 12.5 percent are unconstrained forward-looking individuals. With these fractions, the optimal eligibility age is 61.[29]

[29] Although the actual spike in retirement at age 62 is much smaller than would occur if the population was half myopes, we think of myopia more broadly as reflecting people who retire both at the EEA and at their age of initial private pension entitlement -- so the fraction of people who retire "too early" could be much greater than the spike at age 62.

The two panels on the right of figure 7 repeat this exercise under conditions designed to replicate those of 1962 – the age men first became eligible to receive benefits at age 62. Specifically, we reduce wage levels and initial asset levels by 49 percent to reflect the lower productivity levels that prevailed 40 years ago. We also shift the initial health distribution, the health transition matrix, and the age term in the regression to reflect an 8-year improvement in health as suggested by the analysis in section 4. Finally, we use mortality tables for people born in 1900 rather than 1940. In this alternative simulation, the typical desired retirement age of myopes shifts back by 4 years. The bottom right panel shows that with 50 percent myopes, the optimal EEA is therefore, 57, 4 years earlier than under today's conditions. A clear implication of these findings is that to justify an EEA, much less a rising one, there must be a very large share of the population that retires too early – otherwise concern for the well-being of the constrained forward-looking individuals would suggest that a low EEA is in order.

Table 2 shows the optimal eligibility age from six policy simulations, including the two we have just seen. All of the simulations assume that 50 percent of the population is myopic. The first row presents estimates with γ=1.26. With this value of γ, the AA curve in Figure 1 hardly shifts at all and thus desired retirement ages unambiguously rise over time. The first and third columns correspond to the results we saw in figure 7 – with a four year difference in optimal eligiblity ages between the two simulations. The second column assumes that all of the improvement in health operates through the health coefficient. In this case, the effect of improved health is much smaller, and the difference between 2000 conditions and 1962 conditions shifts the optimal eligibility age by only one year.

The second row of table 2 shows estimates with γ=2. With this value of γ, the AA curve in Figure 1 shifts to the left as the income effect of rising productivity levels exceeds the substitution effect by enough to outweigh the longevity effect. With the AA curve and BB curve going in opposite directions, the shift in the optimal EEA is smaller in these simulations – two years with both health and age are shifted and zero years when only health is shifted.

7. Conclusions

This paper has suggested two models for thinking about the optimal choice of an eligibility age for Social Security. The first balances the need to protect myopic individuals from retiring too early with concerns about forward-looking individuals with low asset levels who would like to start consuming their Social Security benefits at a relatively young age. The second views the Social Security retirement age as the age at which it no longer pays to screen people for disability.

We are currently working in three extensions to the work presented in this paper. First, we are in the process of applying for the restricted version of the HRS so as to incorporate better estimates of lifetime earnings and expected pension benefits into our empirical simulations of the myopia model. Second, we are attempting to calibrate our disability insurance model to real data. Third, we are doing empirical analysis of how the correlation between health status and asset levels on the verge of retirement has changed over time. If the correlation between poor health and low asset levels is rising, then the case for raising the EEA is weakened.

References

Aaron, Henry J. and Robert D. Reischauer (2001). *Countdown to Reform: The Great Social Security Debate* (Washington, DC: Century Foundation).

Bloom, David E., David Canning, and Michael Moore (2004). "The Effect of Improvements in Health and Longevity on Optimal Retirement and Saving," Working paper 10919 (Cambridge, MA: National Bureau of Economic Research).

Cohen, Wilbur J (1985). "The Social Security Act of 1935: Reflections Fifty Years Later," in *The Report of the Committee on Economic Security of 1935 and Other Basic Documents Relating to the Development of the Social Security Act* (Washington, DC: National Conference on Social Welfare).

Coile Courtney C. and Jonathan Gruber (2001). "Social Security Incentives for Retirement," in *Themes in the Economics of Aging*, David A. Wise, ed. (Chicago: University of Chicago Press).

Coile, Courtney C., Peter Diamond, Jonathan Gruber, and Alain Jousten (2002). "Delays in Claiming Social Security Benefits." *Journal of Public Economics* 84(3): 357-385.

Costa, Dora (1998). *The Evolution of Retirement: An American Economic History* 1880-1990. (Chicago: University of Chicago Press).

Cutler, David (2001). "Declining Disability Among The Elderly," *Health Affairs*, 20(6), 11-27.

Cutler, David and Elizabeth Richardson (1997). "Measuring the Health of the United States Population," *Brookings Papers on Economic Activity*, Microeconomics, 1997, 217-272.

Diamond, Peter A. and James A. Mirrlees (1978). "A Model of Social Insurance with Variable Retirement." *Journal of Public Economics* 10: 295-336.

Diamond, Peter A. and James A. Mirrlees (1986). "Payroll-Tax Financed Social Insurance with Variable Retirement." *Scandinavian Journal of Economics* 88: 25-50.

Diamond, Peter A. and Eytan Sheshenski (1995). "Economic Aspects of Optimal Disability Benefits," *Journal of Public Economics* 57: 1-23.

Diamond, Peter and Peter Orszag (2004). *Saving Social Security: A Balanced Approach*. (Washington, DC: Brookings Institution).

French, Eric (2005). "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior," *Review of Economic Studies*, 72: 395-427.

Gustman, Alan L. and Steinmeier, Thomas L. (2005). "The Social Security Early Retirement Age in A Structural Model of Retirement and Wealth." *Journal of Public Economics* 89: 441-463.

Idler, Ellen L. and Stanislav Kasl (1991). "Health Perceptions and Survival: Do Global Evaluations of Health Status Really Predict Mortality?" *Journals of Gerontology* 46: S55-S65.

Meriam, Lewis and Karl Schlotterbeck (1950). *The Cost and Financing of Social Security* (Washington, DC: The Brookings Institution).

Rodgers, Daniel T. (1998). *Atlantic Crossings: Social Politics in a Progressive Age* (Cambridge, MA and London: Belknap Press 1998).

Samwick, Andrew (1998). "New Evidence on Pensions, Social Security, and the Timing of Retirement." *Journal of Public Economics* 70: 207-236.

Sheshinski, Eytan (2003). "Annuities and Retirement," in *Assets, Beliefs, and Equilibria in Economic Dynamics*, in D. Aliprantis, K. Arrow , P. Hammond , F. Kubler, H-M Wu and N. Yannelis (eds). Springer 27-54.

Waidman, Timothy, John Bound, and Austin Nichols (2003). "Disability Benefits as Social Insurance: Tradeoffs Between Screening Stringency and Benefit Generosity in Optimal Program Design." Working paper 042 (University of Michigan: Michigan Retirement Research Center).

Figure 1

The Retirement Decision

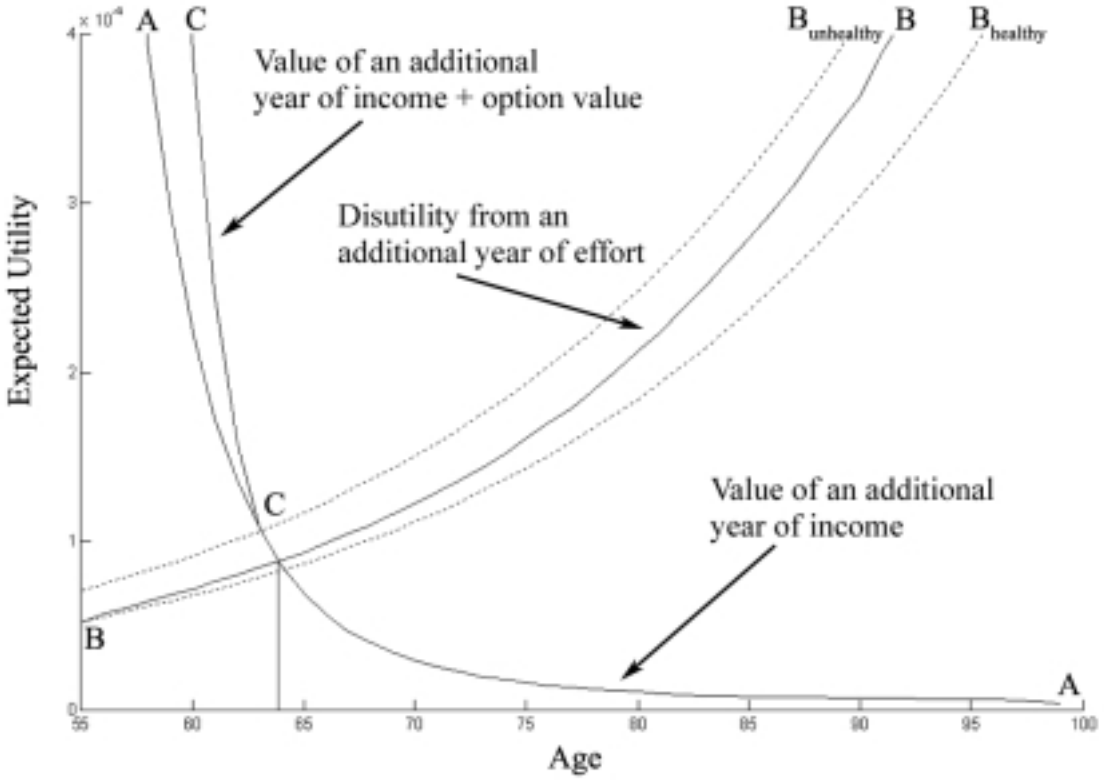


Figure 2
Effects of Improvements in Productivity, Longevity, and Health

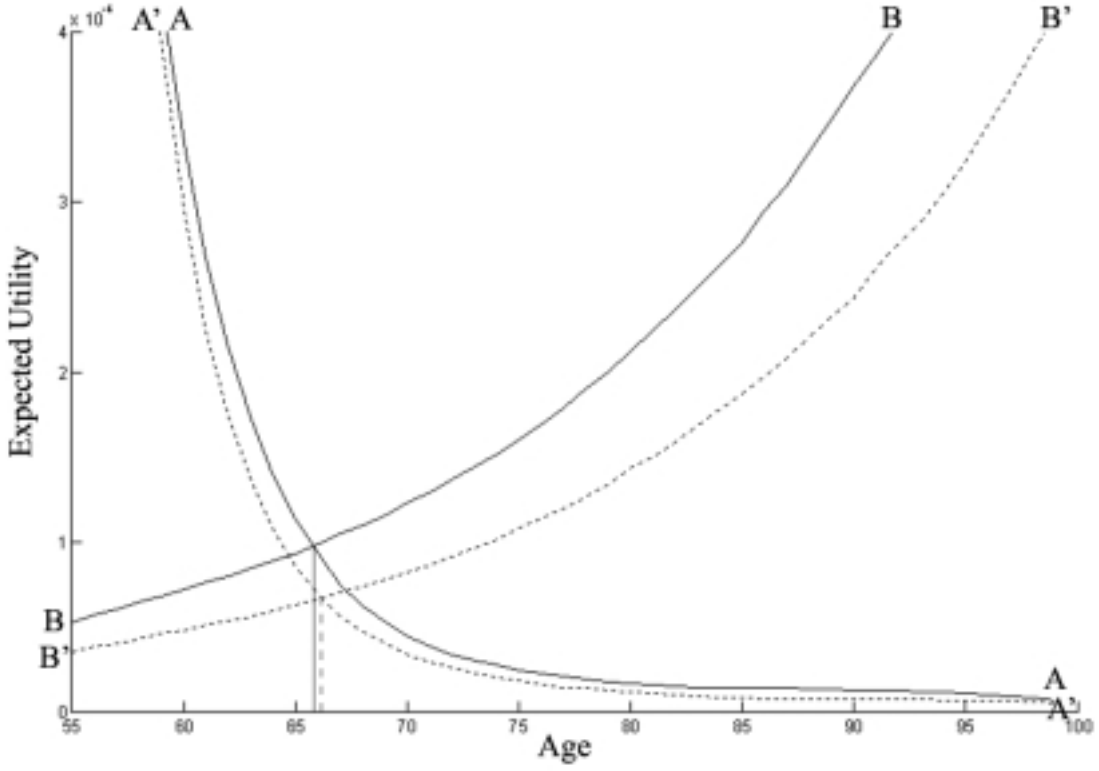
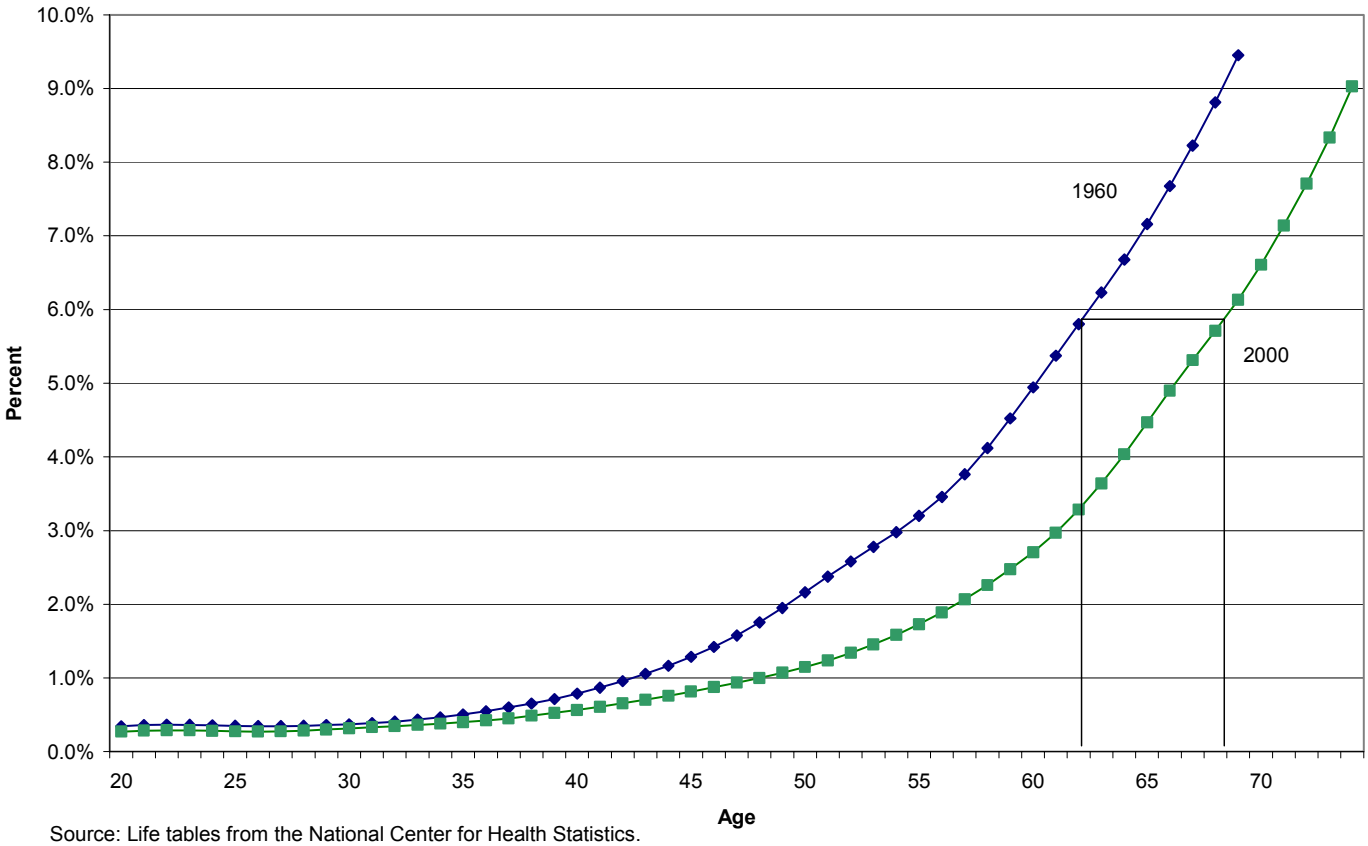


Figure 3: Share of Men in the Last Two Years of Life



Source: Life tables from the National Center for Health Statistics.

Figure 4: Share of Men in Fair or Poor Health

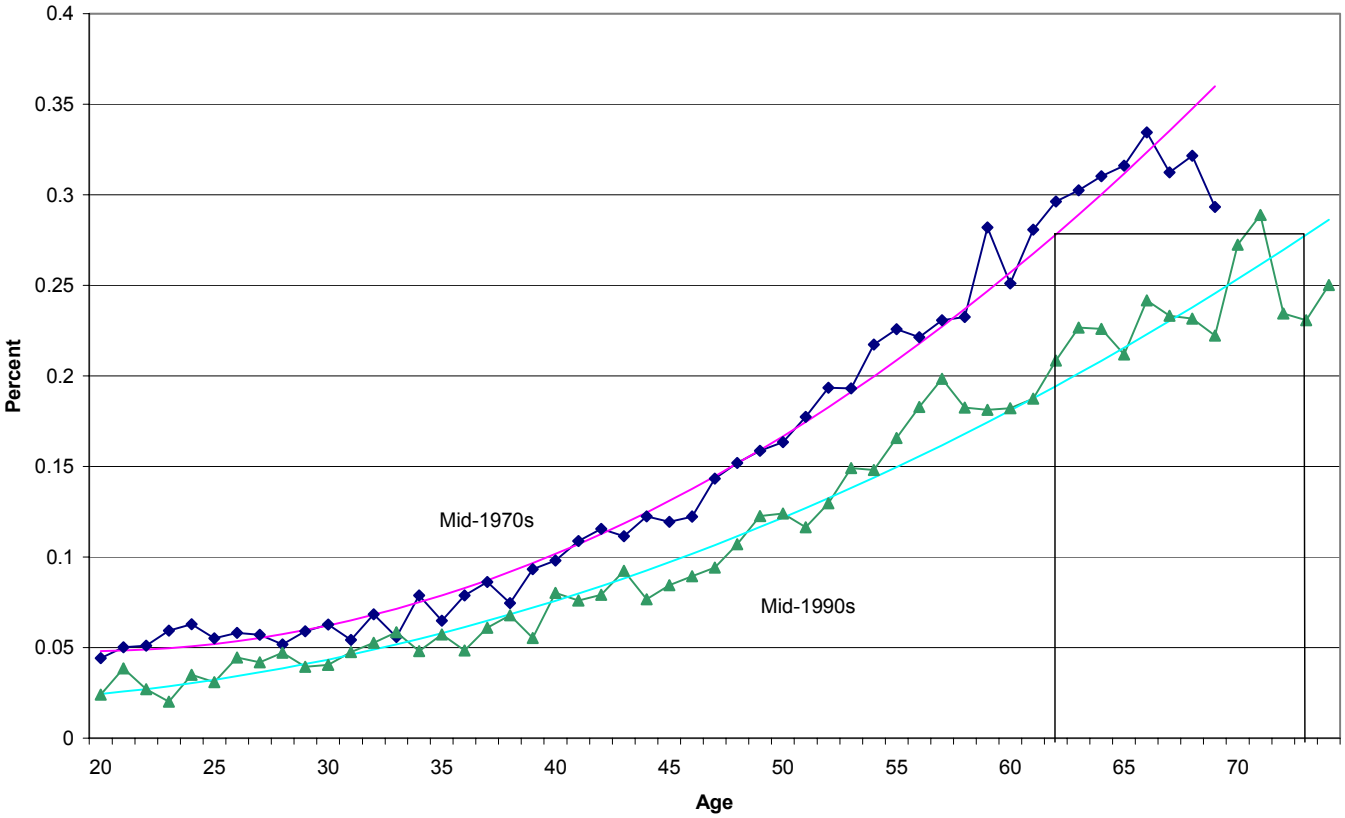


Figure 5: Beds Days in the Past Year for People with Heart Disease

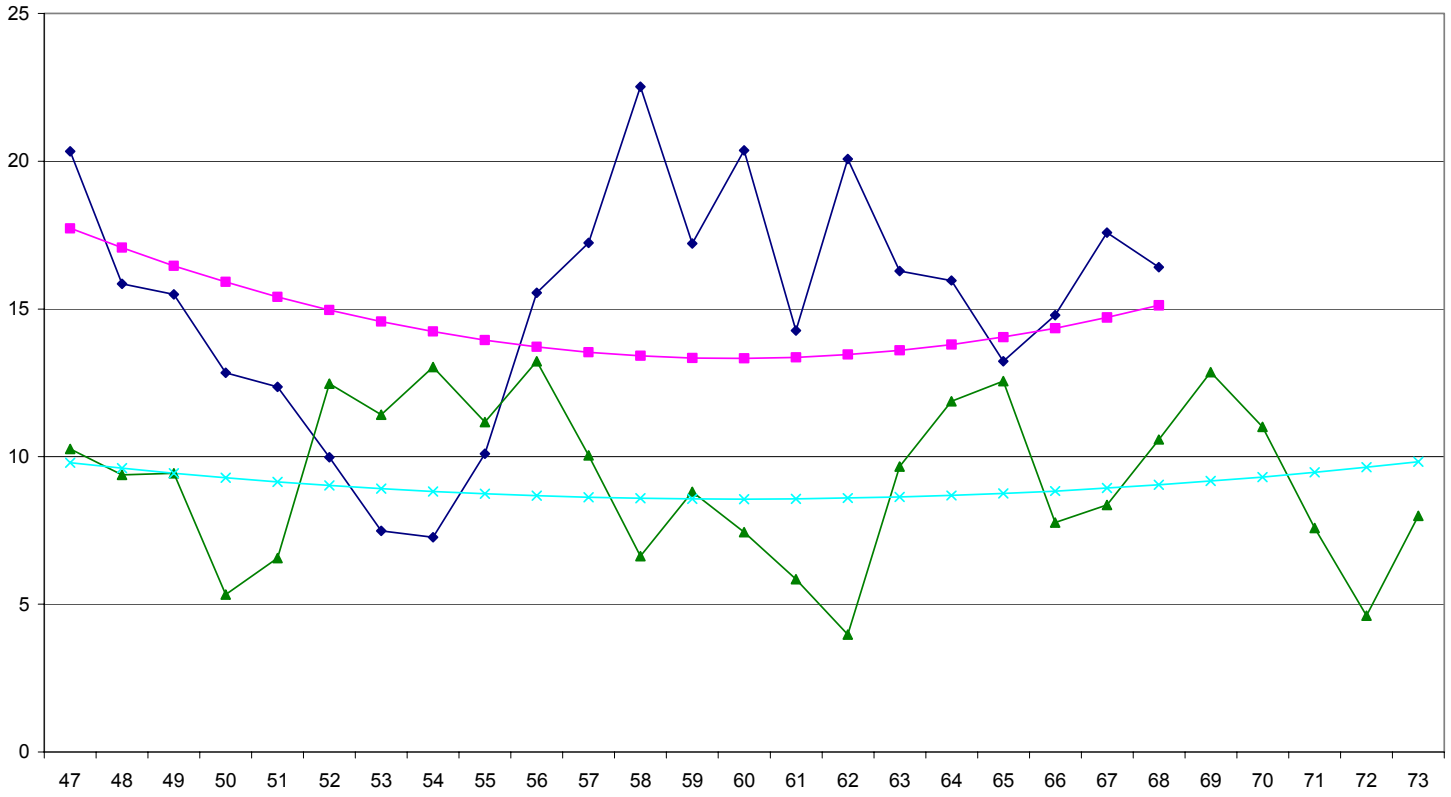


Figure 6
True and Estimated Moments

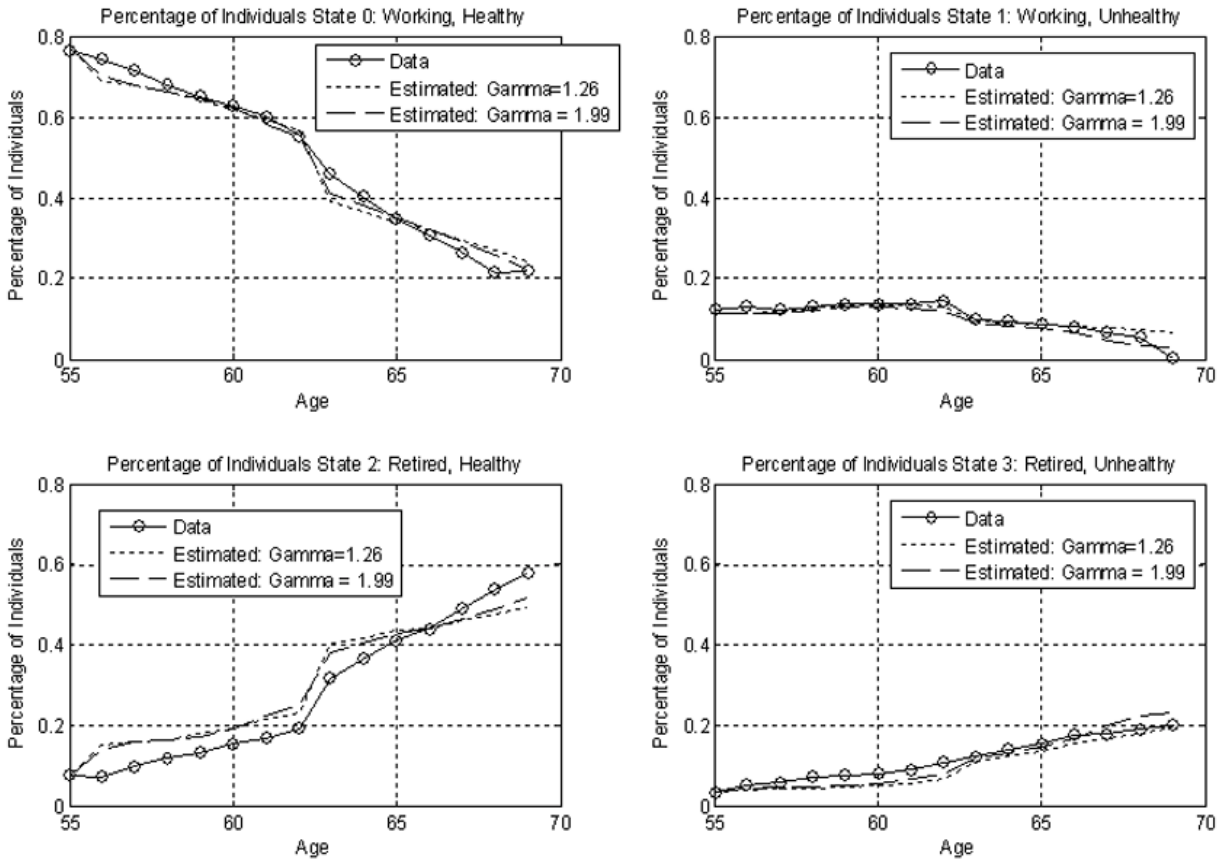
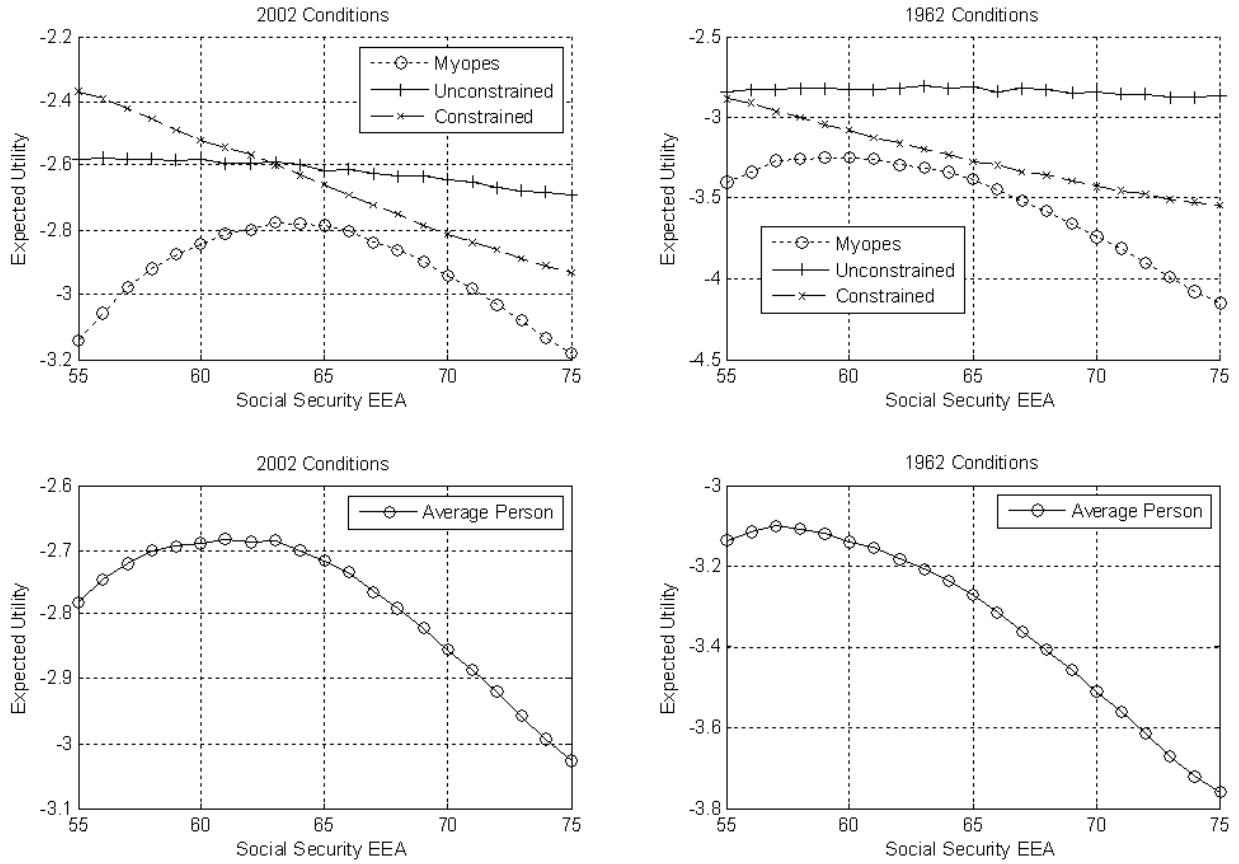


Figure 7
Policy Simulations of Myopia Model



Notes: The simulations on the left represent current conditions. The simulations on the right adjust wages, asset levels, and mortality to 1962 levels and shift both the health and age terms in the disutility of effort by eight years.

Table 1
Estimated Parameters of Retirement Model

Model	β_0	$\beta_{poorhealth}$	β_{age}	σ_e	# of obs
γ=1.26	-3.95	1.29	.100	.577	132
γ=2.00	-13.24	3.92	.125	.012	132
G-S (γ=1.26)	-2.45	.298	.076	.440	

Note: G-S are the results from Gustman and Steinmeier (2005), table 1.

Table 2
Simulations of Optimal Social Security Eligibility Age from Myopia Model

Model	2000	1962 health shifted	1962 health and age shifted
γ=1.26	61	60	57
γ=2.00	62	62	60