THE EFFECTS OF HEALTH INSURANCE AND SELF-INSURANCE ON RETIREMENT BEHAVIOR

Eric French
Federal Reserve Bank of Chicago

John Bailey Jones
SUNY-Albany

November 18, 2002

*Comments welcome at efrench@frbchi.org and jbjoness@csc.albany.edu. We thank Joe Altonji, David Blau, Chris Carroll, Tim Erikson, Hannning Fang, John Kenney, Guy Laroque, Brigitte Madrian, John Rust, Dan Sullivan, and seminar participants at Duke, Chicago Business School, Concordia, UCL, the Board of Governors, the Minneapolis Fed, BLS, INSEE-CREST, the Econometric Society, SED, NBER, the Conference on Social Insurance and Pensions Research, the Society for Computational Economics, and the Federal Reserve System Committee Meetings for helpful comments. Diwakar Choubey, Ken Housinger, Kirti Kamboj, and Santadarshan Sadhu provided excellent research assistance. The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) to the Center for Retirement Research at Boston College. The views of the authors do not necessarily reflect those of the Federal Reserve System or the SSA. Recent versions of the paper can be obtained at http://www.chicagofed.org/economists/EricFrench.cfm/.
Abstract

Using an estimable dynamic programming model, we assess the relative importance of Medicare and Social Security in determining retirement behavior. A key issue is whether individuals value health insurance benefits only because health insurance reduces average medical expenses, or also because it reduces the volatility of expenses. To address this issue, the model explicitly includes volatile medical expenses and volatility-reducing health insurance. Moreover, the model includes a savings decision, so that individuals can self-insure against health cost shocks through saving. Using data from the Health and Retirement Survey, we find that the reduction in expected medical expenses explains 75% of a typical individual’s valuation of health insurance, with the reduction in volatility explaining the remaining 25%. We find that shifting the Medicare eligibility age to 67 will delay age of retirement, but that shifting the Social Security normal retirement age to 67 will cause an even larger delay.
1 Introduction

Prior to age 65, many individuals receive health insurance only if they continue to work. At age 65, however, Medicare provides health insurance to almost everyone. Therefore, a potentially important work incentive disappears at age 65.\(^1\) If individuals place a high value on health insurance, the provision of Medicare benefits may have a large effect on retirement behavior. To see if this is the case, we construct and estimate a retirement model that includes health insurance, uncertain medical costs, a savings decision, a non-negativity constraint on assets and a government-provided consumption floor.

Because the future solvency of the Medicare and Social Security systems depends critically on future labor supply patterns, policymakers are interested in the relative roles of Medicare and Social Security in determining labor supply behavior. Although work hours drop sharply between ages 64 and 65, it is not clear whether Social Security or Medicare is primarily responsible for this drop; both Medicare and Social Security generate work disincentives at age 65.\(^2\) As result, it is not clear whether the current plan to raise the normal Social Security Retirement age from 65 to 67 while keeping the Medicare eligibility age constant at 65 will significantly affect the labor supply of older individuals.

Structural studies of retirement behavior have arrived at different conclusions about the relative importance of Social Security and Medicare. The different conclusions seem to result from differences in how these studies treat market incompleteness and uncertainty, which affect how much individuals value Medicare.

Assuming that individuals value health insurance at the cost paid by employers, both Lumsdaine et al. (1994) and Gustman and Steinmeier (1994) find that health insurance has a small effect on retirement behavior. Their results are potentially driven by the fact that

---

\(^1\)Employer-provided insurance becomes the secondary payer of medical expenses for many retired individuals, so there is still some value to employer-provided insurance after age 65.

\(^2\)Social Security induces retirement at age 65 because there are strong actuarial incentives to begin drawing Social Security benefits by age 65. Benefit receipt in turn discourages work, because the Social Security Earnings Test effectively taxes labor income above a certain level. A fuller description of these incentives appears in Section 4.3.
Social Security accrual rates are actuarially unfair after age 65, while on average the dollar value of employer contributions to health insurance declines by a relatively small amount after age 65. Diamond and Gruber (1997) find that for an average person the work disincentives caused by Social Security jump from about $500 per year at age 64 to $2,500 per year at age 65. In contrast, Gustman and Steinmeier (1994) find that the average employer contribution to employee health insurance is about $2,500 per year before age 65.\(^3\) While the employer contribution declines after age 65, the decline is too small to generate strong retirement incentives. In short, if health insurance is valued at the cost paid for by employers, the age-65 work disincentives of Social Security are much greater than the work disincentives of Medicare.

If individuals are risk-averse, however, and large out-of-pocket Medical costs are possible, individuals could value health insurance well beyond the cost paid by employers.\(^4\) If individuals are uninsured, they will potentially face volatile medical expenses. Volatile medical expenses in turn will lead to volatile life-cycle consumption paths. If individuals are risk-averse, they will value the consumption smoothing that health insurance provides. Therefore, Medicare’s age-65 work disincentive comes not only from the reduction in average medical costs paid by those without employer-provided health insurance, but from also the reduction in the volatility of those costs.

Addressing this point, Rust and Phelan (1997) estimate a dynamic programming model that accounts explicitly for risk aversion and uncertainty about out-of-pocket medical expenses. They find that because of health cost uncertainty, Medicare has large effects on retirement behavior. However, Rust and Phelan assume that an individual’s consumption equals his income net of out-of-pocket medical expenses. In other words, Rust and Phelan ignore an individual’s ability to self-insure against out-of-pocket medical expenses through...

\(^3\)This is taken from the 1977 NMES, then adjusted to 1998 dollars using the medical component of the CPI.

\(^4\)While individuals can usually buy private health insurance, adverse selection problems can make private health insurance prohibitively expensive. For example, Gruber and Madrian (1995) report that one New England insurance company charged $8,640 for family health insurance in 1993, which translates into $10,310 in 1998 dollars. Therefore, in the absence of employer- or government- provided insurance, many people go without health insurance. Employer provided health insurance costs employers on average XX.
saving. Smith (1999) finds that out-of-pocket medical expenses generate large declines in wealth. Cochrane (1991) finds that short-term illnesses generate only small declines in food consumption. These findings suggest that savings might be important. To the extent that Rust and Phelan overstate the amount of consumption volatility caused by out-of-pocket medical cost volatility, they overstate the value of health insurance and thus the effect of health insurance on retirement behavior.

Lumsdaine et al. (1994) and Gustman and Steinmeier (1994) potentially underestimate the value of health insurance, while Rust and Phelan (1997) potentially overestimate it. A major goal of this paper, therefore, is to better assess the value of health insurance. We do this by constructing a model of retirement behavior that not only accounts explicitly for health cost uncertainty and health insurance, but also has a savings decision. This allows us to consider whether uncertainty and self-insurance greatly affects the value of health insurance.

The paper proceeds as follows. Section 2 develops a dynamic programming model of retirement behavior in which individuals face wage and medical expense shocks. Both the mean and the variance of medical expenses depend upon an individual's health insurance status, health status, age, and whether or not the individual has applied for Social Security benefits.

Section 3 describes how we estimate the model using the Method of Simulated Moments (MSM). An interesting feature of the estimation approach is that in addition to matching mean hours and participation at each age, we match multiple asset quantiles, as well as asset-quantile-conditional labor force participation rates. This forces the model to replicate not only the average behavior of all individuals, but to also replicate differences between high and low-wealth individuals, who differ in their ability to self-insure against health cost shocks.

Section 4 describes the data from the Health and Retirement Survey (HRS) that we use in our analysis. We find that those with employer-provided health insurance have, on average, lower and less variable medical expenses than those without employer-provided insurance at

---

5While not unprecedented (see, for example, the minimum distance estimator constructed by Epple and Seig, 1999), this aspect of our approach is fairly novel.
all ages. However the differences in medical expenses are larger before age 65, when Medicare is not available. Therefore, there does exist a drop in work incentives at age 65.

Section 5 presents life cycle profiles drawn from the data. We estimate job exit rates by age, health insurance type and asset level. If access to employer-provided health insurance affects retirement, it should be the case that those who lose their health insurance when they leave their job are more likely to wait until age 65—when they can receive Medicare instead—to leave their jobs. The data, however, show that job exit rates differ little across health insurances. Moreover, we would expect that low-asset individuals, those least able to buffer themselves against out-of-pocket medical expenses, value employer-provided insurance more than high-asset individuals. We find, however, that there do not seem to be large systematic differences in job exit rates by asset level.

Section 5 also contains preliminary preference parameter estimates for the structural model. Section 6 presents preliminary results from an experiment with the model. We measure the changes in lifetime hours of work induced by raising the normal Social Security retirement age to 67, first in isolation and then jointly with the Medicare eligibility age. We find that the incremental effect of raising the Social Security retirement age in isolation is much bigger than the incremental effect of raising the Medicare eligibility age. In order to understand why Social Security is more important than Medicare, we then evaluate how much individuals value health insurance. We find that about 75% of the value of health insurance comes from the reduction in average medical expenses, with the remaining 25% coming from the reduction in medical expense uncertainty. This suggests that self-insurance significantly reduces the effects of health cost uncertainty.

Section 7 concludes.

---

6Government-mandated COBRA coverage extends health insurance coverage to everyone who leaves their job for a short period of time. We account for this in the model.
2 The Model

Consider a household head seeking to maximize his expected lifetime utility at age (or equivalently, year) \( t, t = 1, 2, \ldots, T + 1 \). Each period that he lives, the individual receives utility, \( U_t \), from consumption, \( C_t \), hours worked, \( H_t \), and health (or medical) status, \( M_t \), so that \( U_t = U(C_t, H_t, M_t) \). When he dies, he values bequests of assets, \( A_t \), according to a bequest function \( b(A_t) \). Let \( s_t \) denote the probability of being alive at age \( t \) conditional on being alive at age \( t - 1 \), and let \( S(j, t) = (1/s_t) \prod_{k=t}^{j} s_k \) denote the probability of living to age \( j \), conditional on being alive at age \( t \). Since age \( T + 1 \) is the terminal period, \( s_{T+1} = 0 \). We assume that preferences take the form:

\[
U(C_t, H_t, M_t) + E_t \left[ \sum_{j=t+1}^{T+1} \beta^j S(j-1, t) \left( s_j U(C_j, H_j, M_j) + (1-s_j) b(A_j) \right) \right], \tag{1}
\]

where \( \beta \) is the time discount factor. In addition to choosing hours and consumption, eligible individuals can choose whether to apply for Social Security benefits; let the indicator variable \( B_t \in \{0, 1\} \) equal one if the individual has applied for benefits. The individual maximizes equation (1) by choosing the contingency plans \( \{C_j, H_j, B_j\}_{j=t}^{T+1} \), subject to the constraints on \( \{C_j, H_j, B_j\} \) described in equations (4) through (15) below.

The within-period utility function is of the form

\[
U(C_t, H_t, M_t) = \frac{1}{1-\nu} \left( C_t^\gamma (L - H_t - \theta_P P_t - \phi \times 1\{M_t = \text{bad}\}) \right)^{1-\gamma}, \tag{2}
\]

where the per-year time endowment is \( L \) and the quantity of leisure consumed is \( L - H_t - \theta_P P_t - \phi \times 1\{M_t = \text{bad}\} \). The individual’s utility from leisure depends on his health status through the 0-1 indicator \( 1\{M_t = \text{bad}\} \), which equals one when his health is bad. Participation in the labor force is denoted by \( P_t \), a 0-1 indicator equal to 0 when hours worked, \( H_t \), equals zero. The fixed cost of work, \( \theta_P \), is measured in hours worked per year.\footnote{Fixed costs are frequently cited as a reason why annual hours of work are clustered around both 2000 hours and 0 hours of work (Cogan, 1981).} We treat retirement
as a form of the participation decision, and thus allow retired workers to reenter the labor force.

We assume that the parameter $\gamma$ is between 0 and 1 and the parameter $\nu$ is positive. $\nu$, the coefficient of relative risk aversion for total utility, has two functions. First, it controls the intertemporal substitutability of consumption and leisure. As $\nu$ increases, individuals are less willing to intertemporally substitute. Second, $\nu$ measures the non-separability between consumption and leisure. Under certainty and interiority, $\nu > 1$ implies that leisure and consumption are substitutes (Low, 2000).

The bequest function takes the form

$$b(A_t) = \theta B \frac{(A_t + K)^{(1-\nu)\gamma}}{1 - \nu},$$

where $K$ determines the curvature of the bequest function.

Individuals face several sources of uncertainty. The individual’s health status at time $t$, $M_t$, is a Markov process taking two values, good and bad. We assume that the transition probabilities for health status depend on current health status and age, so that the elements of the health status transition matrix are\(^8\)

$$\pi_{ij}(t) = \Pr(M_{t+1} = j | M_t = i, t), \quad i, j \in \{\text{good, bad}\}.$$  \hspace{1cm} (4)

Mortality rates depend upon age and previous health status:

$$s_{t+1} = s(M_t, t + 1).$$ \hspace{1cm} (5)

We assume that the logarithm of wages\(^9\) at time $t$, $\ln W_t$, is a function of health status,

\(^8\)We ignore the possibility, pointed out by Grossman (1972), that wealth may affect health.

\(^9\)“Wages” include both the observed wage of labor market participants and the potential wage of non-participants.
age, $W(M_t, t)$, hours worked $H_t$ and an autoregressive component, $\omega_t$:

$$\ln W_t = W(M_t, t) + \alpha \ln H_t + \omega_t,$$  \hspace{1cm} \text{(6)}

The inclusion of hours, $H_t$ the wage determination equation captures the empirical regularity that, all else equal, part-time workers earn relatively lower wages than full time workers.\footnote{Among other things, this might reflect fixed costs on the side of the employer.} The autoregressive component $\omega_t$ has the correlation coefficient $\rho_W$ and the normally-distributed innovation $\eta_t$:

$$\omega_t = \rho_W \omega_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma^2_\eta).$$  \hspace{1cm} \text{(7)}

In the interest of computational simplicity, we assume that spousal income is a deterministic function of the individual’s wage and age.\footnote{Note, however, that spouse’s income only depends on the exogenous component of wages, not hours choices.}

$$y_s t = y_s(W(M_t, t) + \omega_t, t).$$  \hspace{1cm} \text{(8)}

Medical expenses, $hc_t$, which are the focus of this paper, are defined as the sum of out-of-pocket costs and insurance premia. We assume that health costs depend upon an individual’s health insurance status ($HI_t$), health status ($M_t$), age ($t$), whether or not the individual has applied for Social Security benefits ($B_t$), and a person-specific effect ($\psi_t$):

$$\ln hc_t = hc(M_t, HI_t, t, B_t) + \sigma(M_t, HI_t, t, B_t) \times \psi_t.$$ \hspace{1cm} \text{(9)}

Note that health insurance affects both the expectation of medical expenses, through $hc(,)$ and the variance, through $\sigma(,)$. These differences across health insurance types usually shrink at age 65, when Medicare becomes the primary insurer of Social Security recipients.

\footnote{This follows most studies of wage dynamics (Farber and Gibbons, 1996, French, 1999) which find that wages follow a highly persistent AR(1) process.}
Following Feenberg and Skinner (1994) and French and Jones (2002), we assume that the idiosyncratic component of medical expenses $\psi_t$ can be decomposed as

\begin{align*}
\psi_t & = \zeta_t + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2), \\
\zeta_t & = \rho_{he}\zeta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2),
\end{align*}

where $\xi_t$ and $\epsilon_t$ are serially and mutually independent. $\xi_t$ is the transitory component of health cost uncertainty, while $\zeta_t$ is the persistent component, with autocorrelation $\rho_{he}$. As a matter of timing, we assume that the time-$t$ shocks to medical expenses, $\xi_t$ and $\epsilon_t$, are not observed until period $t$ has ended, so that time-$t$ shocks do not affect time-$t$ labor decisions. We view this assumption as more reasonable than the alternative, namely that the time-$t$ shocks are fully known when workers decide whether to hold on to their employer-provided health insurance.

Differences in labor supply behavior across categories of health insurance coverage are an important part of identifying our model. We assume that there are four mutually exclusive health insurance categories. The first is retiree coverage, $ret$, where workers keep their health insurance even after they leave their job. The second category is tied health insurance, $tied$, where workers receive employer-provided coverage as long as they continue work. If a worker with tied health insurance leaves his job, however, he receives “COBRA” coverage, COBRA, for one year, and then his insurance ceases.\footnote{Although there is some variability across states as to how long individuals are eligible for COBRA coverage, by Federal law most individuals are covered for 18 months (Gruber and Madrian, 1995). Given a model period of one year, we approximate the 18-month period as a one-year term.} The fourth category consists of individuals whose potential employers provide no health insurance at all, or $none$.\footnote{Workers in the $none$ category buy insurance on their own, receive some sort of government coverage, or simply go uncovered. We assume that all three groups have the same mean and variance of medical expenses (before the consumption floor, described below).} Individuals that are age 65 or older and have applied for Social Security benefits are also eligible for Medicare health insurance. In other words, those whose health insurance status is $none$ and who are older than 65 are eligible for Medicare. We assume that workers move between these health
status categories according to:  

\[ HI_t = \begin{cases} 
  ret & \text{if } HI_{t-1} = ret \\
  tied & \text{if } HI_{t-1} = tied \text{ and } H_t > 0 \\
  COBRA & \text{if } HI_{t-1} = tied \text{ and } H_t = 0 \\
  none & \text{if } HI_{t-1} = none \text{ or } HI_{t-1} = COBRA
\end{cases} \tag{12} \]

The individual’s non-pension, non-Social Security assets obey the following accumulation equation:

\[ A_{t+1} = A_t + Y(rA_t + W_tH_t + y s_t + pb_t, \tau) + ss_t + tr_t - h c_t - C_t, \tag{13} \]

where \( Y(rA_t + W_tH_t + y s_t + pb_t, \tau) \) denotes post-tax income, \( r \) denotes the interest rate, \( pb_t \) denotes pension benefits, \( \tau \) is a vector describing the tax structure (described in Appendix A), \( ss_t \) denotes Social Security benefits, and \( tr_t \) denotes government transfers.

Pension benefits are a function of the worker’s age and pension wealth. Pension wealth in turn depends on pension accruals, which are themselves a function of a worker’s age and labor income. The pension accrual formula captures the fact that high income workers have greater pension accrual rates than low income workers and that pension accrual rates are greater for workers in their 50s than in other ages. Details of how we calculate pension accrual, benefits and wealth are in Appendix B. When finding an individual’s decision rules (see Appendix C), we further assume that the individual’s pension wealth can be expressed as a function of his Social Security wealth and age. Details of this are in Appendix B as well.

The variable \( ss_t \) denotes Social Security benefits net of the Earnings Test. Benefits are zero until the individual has applied for Social Security benefits, i.e. until \( B_t = 1 \). An individual must be at least 62 years old to apply. Upon applying an individual receives

\footnote{In imposing this transition rule, we are assuming that people out of the work force are never offered jobs with insurance coverage, and that workers with tied coverage never upgrade to retiree coverage. Restricting an individual’s insurance options in this way most likely leads us to overstate the value of employer-provided health insurance. Given that we find rather small effects of employer-provided health insurance on retirement, adjusting for these self-insurance options would lead us to find even smaller effects.}
benefits until death, i.e. $B_{t+1} = 1$ if $B_t = 1$. Once an individual has applied, his Social Security benefits depend on his Average Indexed Monthly Earnings, or $AIME_t$, which is roughly the 35 highest earnings years in the labor market. Section 4.3 describes specifics of how $AIME_t$, age of application for benefits, and labor income affect $ss_t$. Appendix D describes the computation of $AIME_t$.

Associated with the budget rule is a borrowing constraint:

$$A_t + Y_t + ss_t + tr_t - C_t \geq 0, \quad (14)$$

Because it is illegal to borrow against Social Security benefits and difficult to borrow against most forms of pension wealth, individuals with low non-pension, non-Social Security wealth may not be able to finance their retirement before their Social Security benefits become available, at age 62.

Following Hubbard et al. (1994, 1995), we assume that government transfers provide a consumption floor:

$$tr_t = \max\{0, C_{\min} - (A_t + Y_t + ss_t)\}, \quad (15)$$

Equation (15) says that government transfers bridge the gap between an individual’s “liquid resources” (the quantity in the inner parentheses) and the consumption floor. Equation (15) also implies that if transfers are positive, $C_t = C_{\min}$. Our treatment of government transfers implies that individuals can always consume at least $C_{\min}$, even if their out-of-pocket medical expenses exceed their financial resources. Since the government is effectively providing low-asset individuals with health insurance, these people may place a low value on employer-provided health insurance. This of course depends on the value of $C_{\min}$; if $C_{\min}$ is low enough, it will be the low-asset individuals who value health insurance most highly. Those with very high asset levels should be able to self-insure.

---

16Given the timing of medical expenses described above, under this borrowing constraint an individual with extremely high medical expenses this year could have negative net worth next year. Given that many people in our data still have unresolved medical expenses, medical expense debt seems reasonable.
In recursive form, the individual’s problem can be written as

\[ V_t(X_t) = \max_{C_t, H_t, B_t} \left\{ \frac{1}{1 - \nu} \left( C_t^\gamma (L - H_t - \theta_P P_t - \phi \times 1\{M_t = bad\})^{1-\gamma} + \beta(1 - s_{t+1}) b(A_{t+1}) + \beta s_{t+1} \int V_{t+1}(X_{t+1}) dF(X_{t+1} | X_t, t, C_t, H_t, B_t) \right\}, \]

subject to equations (14) and (15). The vector \( X_t = (A_t, B_t, M_t, \text{ALME}_t, HI_t, \omega_t, \zeta_{t-1}) \) contains the individual’s state variables, while the function \( F(\cdot | \cdot) \) gives the conditional distribution of these state variables.\( ^{17} \) In doing so, \( F(\cdot | \cdot) \) incorporates the budget constraints and stochastic processes described in equations (4) through (13).

An individual’s decisions thus depend on his state variables, \( X_t \), his preferences, \( \theta \), and his beliefs, \( \chi \), where

\[ \theta = (\gamma, \nu, \theta_P, \theta_B, \phi, L, \beta), \]

\[ \chi = (r, w(M_t, H_t, t), \sigma^2_{\eta}, \rho_W, h(c(M_t, H I_t, t, B_t), \sigma(M_t, HI_t, t, B_t), \sigma^2_{\xi}, \sigma^2_{\epsilon}, \rho_{hc}, \{ \text{prob}(M_{t+1} | M_t, t) \}_{t=1}^T, \{ S_t \}_{t=1}^T, \{ Y(\cdot, \cdot) \}, \{ s_{st} \}_{t=1}^T, \{ p_{bt} \}_{t=1}^T, \{ tr_t \}_{t=1}^T). \]

It follows that the solution to the individual’s problem consists of the set of consumption \( \{ C_t(X_t, \theta, \chi) \}_{1 \leq t \leq T} \), work \( \{ H_t(X_t; \theta, \chi) \}_{1 \leq t \leq T} \) and benefit application \( \{ B_t(X_t; \theta, \chi) \}_{1 \leq t \leq T} \) rules that solve equation (16). The labor force participation rule \( P_t(X_t; \theta, \chi) \) is a 0-1 indicator equal to zero when \( H_t(X_t; \theta, \chi) = 0 \). Inserting these decision rules into the asset accumulation equation yields next period’s assets, \( A_{t+1}(X_t, \psi_t; \theta, \chi) \).

Given that the model lacks a closed form solution, these decision rules must be found numerically. Appendix C describes our numerical methodology.

\( ^{17} \)Spousal income and pension benefits (see Appendix B) depend only on the other state variables and are thus not state variables themselves.
3 Estimation

Our goal is to estimate preferences, \( \theta \), and beliefs, \( \chi \). Computational concerns lead us to use a two-step strategy, similar to the ones used by Gourinchas and Parker (2002) and French (2000). In the first step we estimate some belief parameters and calibrate others. In doing this we assume that individuals have rational expectations, so that the belief parameters can be derived by standard statistical methods. We describe the belief parameters in Section 4. In the second step we estimate preference parameters using the method of simulated moments (MSM). In particular, we use the numerical methods described in the previous section and the belief parameters to simulate life cycle profiles. The preference parameters that generate simulated profiles best “matching” the profiles estimated from data are considered to be the true preference parameters. In the next three sections, we describe our MSM methodology in more detail.

3.1 Moment Conditions

The objective of MSM estimation is to find a vector of preferences \( \hat{\theta} \) yielding simulated life-cycle decision profiles that “look like” (as measured by a GMM criterion function) the profiles from the data. A key part of the MSM approach is selecting which features of the data distribution—which profiles—to match. The model predicts that labor supply behavior should differ by age, health insurance status, health status, medical expenses, and asset level. We therefore require our model to match labor force participation conditional on asset grouping and health insurance status. Moreover, because one’s ability to self-insure against medical expense shocks depends critically upon one’s asset level, we match asset quantiles. We also match participation rates and hours by overall health status.

Our motivation for using the MSM approach is best illustrated with actual moment conditions. Consider assets. Letting \( A_{it+1}, X_{it} \) and \( \psi_{it} \) denote, respectively, individual \( i \)'s observed asset level, state vector, and health cost shock, we have

\[
A_{it+1} = A_{t+1}(X_{it}, \psi_{it}; \theta, \chi) + u_{it},
\]
where $u_{it}$ is defined as the remainder. If at the “true” parameter vectors $\theta_0$ and $\chi_0$, $u_{it}$ is quantile-independent of $X_{it}$ and $\psi_{it}$, one can consistently estimate $\theta_0$ using quantile regression.

Quantile independence, however, is a fairly strong restriction that is unlikely to hold. For example, there is probably measurement error in some of the state variables, such as the wage measure. A less demanding approach is to match unconditional quantiles.\(^{18}\) To see how this works, let $j \in \{1, \ldots, J\}$ index quantiles. In practice we consider the 1/3rd and 2/3rd quantiles, so that $J = 2$. Assume that assets, $A_{it}$, have a continuous density. Let $g_{\pi_j}(t; \theta_0, \chi_0)$ denote the value of the $\pi_j$-th asset quantile predicted by the model. This means, for example, that if $\pi_1 = 1/3$ and $g_{\pi_1}(53; \theta_0, \chi_0) = \$50,000$, the model predicts that 1/3 of all individuals have assets of $\$50,000$ or less at age 53. In Appendix E we show that if the model is correctly specified,

$$E \left( 1\{ A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0) \} - \pi_j | t \right) = 0,$$

(17)

for $j \in \{1, 2, \ldots, J\}$, $t \in \{1, \ldots, T\}$. Since $J = 2$, equation (17) generates $2T$ moment conditions.

The key to the MSM approach is that one can derive $g_{\pi_j}(t; \theta_0, \chi_0)$ by finding the model’s decision rules for consumption, hours, and benefit application, using the decision rules to generate artificial histories for many different simulated individuals, and finding the quantiles of the collected histories. Equation (17) therefore says that the data sample and the simulated sample have the same age-conditional asset quantiles. It is worth stressing that the distribution used to derive $g_{\pi_j}(t; \theta_0, \chi_0)$ is found by evaluating $A_{t+1}(X_{it}, \psi_{it}; \theta, \chi)$ over the simulated distributions of the state vector $X_{it}$ and health cost shock $\psi_{it}$, rather than the empirical distributions; some of the state variables are likely mismeasured. This does not imply, however, that $A_{it}$ cannot vary for reasons, such as measurement error, that are not incorporated into the simulations; the only requirement is that these effects on the data “average out,” so that the expectation in equation (17), which is taken over the observed

\(^{18}\)Gourinchas and Parker (2002) develop this point in more detail in the context of a model of lifetime consumption. These problems are possibly even more severe when estimating the supply of labor hours; given that we use earnings divided by hours as the wage measure, measurement error in hours enters into both measured wages and measured hours, creating the well-known “division bias” problem.
data, continues to hold.

Next, consider how a worker’s asset quantile and health insurance status affects his participation. Let \( \mathcal{P}_j(HI, t; \theta_0, \chi_0) \) denote the model-predicted labor force participation rate conditional upon assets being in the \( j \)-th quantile interval and health insurance being of type \( HI \). If the model is correctly specified,

\[
E \left( P_{it} - \mathcal{P}_j(HI, t; \theta_0, \chi_0) \mid HI_{it} = HI, g_{\pi_{j-1}}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0), t \right) = 0, \tag{18}
\]

for \( j \in \{1, 2, ..., J + 1\}, \; HI \in \{\text{none, ret, tied}\}, \; t \in \{1, ..., T\}. \]

Equation (18) says that within each asset grouping, the data sample and the simulated sample have the same conditional mean. With 2 quantiles (generating 3 quantile conditional means) and 3 health insurance types, equation (18) generates \( 9T \) moment conditions.

Finally, consider health-conditional hours and participation. Let \( \mathcal{H}(M, t; \theta_0, \chi_0) \) and \( \mathcal{P}(M, t; \theta_0, \chi_0) \) denote the conditional expectation functions for hours (when working) and participation generated by the model; let \( H_{it} \) and \( P_{it} \) denote measured hours and participation. If the model is correctly specified,

\[
E \left( \ln H_{it} - \ln \mathcal{H}(M, t; \theta_0, \chi_0) \mid P_{it} > 0, M_t = M, t \right) = 0, \tag{19}
\]

\[
E \left( P_{it} - \mathcal{P}(M, t; \theta_0, \chi_0) \mid M_t = M, t \right) = 0, \tag{20}
\]

for \( t \in \{1, ..., T\}, \; M \in \{\text{good, bad}\} \). Equations (19) and (20) yield \( 4T \) moment conditions. Combined with the \( 2T \) moment conditions for the asset quantiles and the \( 9T \) moment conditions for asset- and insurance-conditional participation, this generates a total of \( 15T \) moment conditions.

\footnotetext{Because we are interested in participation given one’s opportunity set, we combine individuals who work and receive \textit{tied} insurance with those who do not work and receive \textit{COBRA} coverage, as those two groups had the same insurance opportunities.}
3.2 Estimation Mechanics

The mechanics of our MSM procedure are as follows. Recall that throughout the procedure we take as given the estimates of the parameter vector $\chi$ described in Section 4. First, we estimate life cycle profiles from the data for hours, participation and assets. Second, using the same data used to estimate the profiles, we generate an initial distribution for health, health insurance status, wages, medical expenses, AIME and assets. We also generate matrices of random health, wage and health cost shocks. The matrices hold shocks for 10,000 simulated individuals over their entire lives. Third, we compute the decision rules for the parameter vector $\theta$, using $\chi$ and the numerical methods described in Appendix C.

The fourth step is to simulate profiles for the decision variables. Each simulated individual receives a draw of assets, health, wages and medical expenses from the initial distribution, and is assigned one of the simulated sequences of health, wage and health cost shocks. With the initial distributions and the sequence of shocks, we then use the decision rules to generate that person’s decisions over the life cycle. Each period’s decisions determine the conditional distribution of next period’s states, and the simulated shocks pin the states down exactly.

Fifth, we aggregate the simulated data into profiles in the same way we aggregated the true data. Sixth, we compute moment conditions, i.e., we find the distance between the simulated and true profiles. Finally, we pick a new value of $\theta$ and repeat the whole process.\(^{20}\) The value of $\theta$ that minimizes the distance between the true data and the simulated data described in equations (17)-(20), $\hat{\theta}$, is the estimated value of $\theta$. We discuss the asymptotic distribution of the parameter estimates, the weighting matrix and the overidentification tests in Appendix E.

3.3 Data Profiles

The MSM algorithm requires us to estimate life cycle profiles for assets, hours, and participation rates. We estimate the profiles using the cross section (i.e. the synthetic cohort).

\(^{20}\)We search over the parameter space $\Theta$ with a simplex algorithm written by Bo Honore and Ekaterini Kyriazidou.
Since the HRS covers a fairly narrow age range, any bias caused by not accounting for cohort effects should be small for the non-wage variables.

4 Data and Calibrations

4.1 HRS Data

We estimate the model with data from the first five waves of the Health and Retirement Survey (HRS). The HRS surveys individuals every two years starting in 1992, so data cover the period 1992-2000. We use men ages 51-69 in the analysis. A description of some of the variables follows.

Hours are the product of usual hours per week and usual weeks per year. To compute hourly wages, the respondent is asked about how they are paid, how often they are paid, and how much they are paid. If the worker is salaried, for example, annual income is the product of pay per period and the number of pay periods per year. The wage is then annual earnings divided by annual hours. If the worker is hourly, we use his reported hourly wage. We treat a worker’s hours for the non-survey (odd) years as missing.

The HRS has a comprehensive asset measure. It includes the value of housing, other real estate, autos, liquid assets (which includes money market accounts, savings accounts, T-bills, etc.), IRAs, stocks, business wealth, bonds, and “other” assets, less the value of debts. For non-survey years, we assume that assets take on the value reported in the preceding year. This implies, for example, that we use the 1992 asset level as a proxy for the 1993 asset level.\textsuperscript{21}

To measure health status we use responses to the question: “would you say that your health is excellent, very good, good, fair, or poor?” If the individual responded fair or poor we consider him in bad health and in good health otherwise. We treat the health status for non-survey years as missing.

\textsuperscript{21}We impute the odd-year asset values, rather than treat them as missing, because we compute labor force participation conditional on both health insurance type and asset level. Given that wealth changes rather slowly over time, these imputations should not severely bias the results.
For survey years the individual is considered in the labor force if he reports working over 300 hours per year. The HRS also asks respondents retrospective questions about their work history. We use the work history to construct a measure of whether the individual worked in the intervening year. For example, if an individual withdraws from the labor force between 1992 and 1994, we use the 1994 interview to infer whether the individual was working in 1993.

4.2 Health Insurance Status and Costs

We assign individuals to one of four mutually exclusive health insurance groups: ret, tied, COBRA, and none, as described in section 2. Because of small sample problems, the none group includes those with no insurance and those with private insurance. Neither type receives employer-provided coverage. Those with private insurance must pay high prices for insurance and those with no insurance most face a high level of medical expense risk. Because the model includes a consumption floor to capture the insurance provided by Medicaid, the none group also includes those whose only form of health insurance is Medicaid. We assign those who have health insurance provided by their spouse to the ret group, along with those who report that they would be able to keep their health insurance after they left their job. Neither of these type have their health insurance tied to their job.

Unfortunately, the HRS has information on health insurance outcomes, not choices. This is an important problem for individuals out of the labor force with no health insurance; it is unclear whether these individuals could have purchased COBRA coverage but elected not to do so. To circumvent this problem we use health insurance in the previous wave and the transitions implied by equation (12) to predict health insurance options. For example, if an individual had health insurance that was tied to his job and was working in the previous wave, that individual’s choice set is tied health insurance and working or COBRA insurance and not working.

\footnote{For example, XX\% of HRS respondents who report having tied health insurance two years before the survey date and report working one year before the survey date report having no health insurance on the survey date.}
Blau and Gilleskie (2000) provide evidence that equation (12) does not list every kind of health insurance transition that actually occurs. For example, they find that many individuals report having retiree coverage one period and tied health insurance coverage the next. It is not clear, however, whether most of these additional transitions reflect actual changes in coverage, or simply measurement error.23

The HRS has data on self-reported medical expenses. Medical expenses are the sum of insurance premia, drug costs, and out of pocket costs for hospital, nursing home care, doctor visits, dental visits, and outpatient care. We fit these data to the health cost model described in Section 2. Because of small sample problems, we allow the conditional mean, \( hc(\cdot) \), and standard deviation, \( \sigma(\cdot) \), to depend only on the individual’s Medicare eligibility and health insurance type. Following the procedure described in French and Jones (2002), \( hc(\cdot) \) and \( \sigma(\cdot) \) are set so that the model replicates the mean and 99.5th percentile of the cross-sectional distribution of medical expenses (in levels, not logs) in each of these categories.

An additional problem is that we observe only what the household pays for medical care. The appropriate measure of medical expenses, however, includes medical expenses paid by the government, because our model explicitly accounts for government transfers. Therefore, we assign Medicaid payments to households that received Medicaid benefits. The 2000 Green Book (Committee on Ways and Means, 2000, p. 923) reports that in 1998 the average Medicaid payment was $10,242 per beneficiary. Starting with this average, we then assume that Medicaid payments have the same volatility as the medical care payments made by uninsured households. This allows us to generate a distribution of Medicaid payments.

Table 1 presents some summary statistics. Table 1 shows that for those younger than 65—and thus ineligible for Medicare—average annual medical costs are $3,200 for those with

---

23In waves 1 and 2 of the HRS those who are coded as having retiree coverage are those who respond “yes” to the question “Is this health insurance available to people who retire?” However, it is not clear whether a “yes” response merely means that they would be eligible for COBRA coverage or whether they would be eligible for retiree coverage. In waves 3, 4, and 5 those who are coded as having retiree coverage are those who respond “yes” to the question “If you left your current employer now, could you continue this health insurance coverage up to the age of 65?” The criteria in waves 3 through 5 is probably a more accurate measure of whether the individual has retiree coverage. Thus the fraction of individuals who reported that they would be unable to have retiree coverage approximately doubled between waves 2 and 3.
retiree coverage and $6,900 for those without employer-provided coverage. Therefore, the average person with no employer-provided health insurance spends $3,700 more per year than the average person with retiree coverage. As Rust and Phelan (1997) and Blau and Gilleskie (2000) argue, it is not just the mean of the medical expenses associated with given health insurance type that determines the insurance’s value, but also their variance and skewness. If health insurance reduces health cost uncertainty, risk-averse individuals may value health insurance at well beyond the cost paid by employers. To give a sense of the volatility, Table 1 also presents the standard deviation and 99.5th percentile of the health cost distributions. Table 1 shows that for those younger than 65, the standard deviation of medical expenses under retiree coverage is $4,400. The corresponding figure for those with no health insurance is $13,600, roughly three times as large. In short, individuals with no health insurance face more health cost risk.

Continuing, Table 1 shows that among individuals that are at least 65 years old—and eligible for Medicare—the average person with no insurance spends only about $1,700 more than the person with retiree coverage. The risk differential shrinks after age 65 as well. For those aged 65 and older, the standard deviation of medical expenses under retiree coverage is $4,400. The corresponding figure for those with no health insurance is $8,000. All of this suggests that the value of employer-provided health coverage declines markedly once individuals become eligible for Medicare.

<table>
<thead>
<tr>
<th>Age</th>
<th>Retiree</th>
<th>Tied</th>
<th>COBRA</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 65</td>
<td>$3,200</td>
<td>$3,100</td>
<td>$4,300</td>
<td>$6,900</td>
</tr>
<tr>
<td>Mean</td>
<td>$4,400</td>
<td>$4,100</td>
<td>$8,300</td>
<td>$13,600</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$26,900</td>
<td>$25,000</td>
<td>$48,900</td>
<td>$79,800</td>
</tr>
<tr>
<td>99.5th Percentile</td>
<td>$26,700</td>
<td>$33,700</td>
<td>$48,600</td>
<td>$48,400</td>
</tr>
<tr>
<td>≥ 65</td>
<td>$3,300</td>
<td>$3,900</td>
<td>$4,400</td>
<td>$5,000</td>
</tr>
<tr>
<td>Mean</td>
<td>$4,400</td>
<td>$5,500</td>
<td>$8,200</td>
<td>$8,000</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$26,700</td>
<td>$33,700</td>
<td>$48,600</td>
<td>$48,400</td>
</tr>
<tr>
<td>99.5th Percentile</td>
<td>$26,700</td>
<td>$33,700</td>
<td>$48,600</td>
<td>$48,400</td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics for Medical expenses, by Age and Health Insurance

24This amount is comparable to EBRI’s (19XX) estimate that employers contribute $3200 on average to employees health insurance.
The parameters for the idiosyncratic process $\psi_t$, $(\sigma^2_\xi, \sigma^2_\epsilon, \rho_{hc})$, were estimated in French and Jones (2002). Table 2 presents the parameters, which have been normalized so that overall variance, $\sigma^2_\psi$, is one. Table 2 reveals that at any point in time, the transitory component generates over 65 percent of the cross-sectional variance in medical expenses. The results in French and Jones (2002) reveal, however, that most of the variance in cumulative lifetime medical expenses is generated by innovations to the persistent component. Given the autocorrelation coefficient $\rho_{hc}$ of 0.954, this is not surprising.

### Table 2: The Variance and Persistence of Innovations to Medical Expenses

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>innovation variance of persistent component</td>
<td>0.03113</td>
</tr>
<tr>
<td>$\rho_{hc}$</td>
<td>autocorrelation of persistent component</td>
<td>0.954</td>
</tr>
<tr>
<td>$\sigma^2_\xi$</td>
<td>innovation variance of transitory component</td>
<td>0.6537</td>
</tr>
</tbody>
</table>

4.3 Social Security

There are three major incentives provided by the Social Security System,\footnote{We use tax and benefit formulas from the Social Security Handbook Annual Statistical Supplement for the year 1998.} each of which tend to induce exit from the labor market when old. First, because Social Security benefits depend upon Average Indexed Monthly Earnings (AIME), which is average earnings in the 35 highest earnings years, increased income when young results in increased Social Security benefits when old. This generates incentives for work tend to offset the payroll tax disincentives for workers during their first 35 years in the labor market. But after 35 years in the labor market, AIME is recomputed upwards only if current earnings exceed earnings in some previous year of work. This means there is a decline in work incentives after 35 years in the labor market. We describe the computation of AIME in more detail in Appendix D.

Second, the age at which the individual applies for Social Security benefits affects the level of benefits. Individuals can first apply for Social Security benefits at age 62. For every year before age 65 the individual applies for benefits, benefits are reduced by 6.7%. This is roughly actuarially fair. But for every year after age 65 that benefit application is delayed,
benefits rise by 6% up until age 70. This is less actuarially fair, and encourages people to apply for benefits by age 65.

Third, the Social Security Earnings Test taxes the labor income of Social Security beneficiaries at a high rate. This reduces post-tax wages for Social Security beneficiaries. When combined with the aforementioned incentives to draw Social Security benefits by age 65, the Earnings Test discourages work after age 65. The Social Security Earnings Test is imposed on beneficiaries younger than age 70. For individuals aged 62-64, each dollar of labor income above the “test” threshold of $9,120 leads to a 1/2 dollar decrease in Social Security benefits, until all benefits have been taxed away. For individuals aged 65-69, any labor income above a threshold of $14,500 leads to a 1/3 dollar decrease in Social Security benefits, until all benefits have been taxed away.

When current benefits are taxed away by the Earnings Test, future benefits are increased. If a year’s worth of benefits are taxed away between 62 and 65, benefits in the future will be increased by 6.7%. If a year’s worth of benefits are taxed away between 65 and 70, benefits in the future will be increased by 6%. While increases in future benefits roughly replace the benefits lost through the Earnings Test for ages 62-64, the benefit adjustments provided in the latter years of the Earnings Test are relatively smaller. Moreover, the Earnings Test tax on benefits is in addition to Federal, state, and payroll taxes on income, so that labor income is taxed twice.

All of these incentives are incorporated in the calculation of Social Security benefits, \( s_{st} \), for asset accumulation equation (13).

### 4.4 Wages

The deterministic part of logged wages is specialized as

\[
\alpha \times \ln(H_t) + w(M_t, t) = 0.415 \times \ln(H_t) + w(M_t, t).
\]  

(21)
The first term on the right, $0.415 \times \ln(H_t)$, is taken from Aaronson and French (2002). This term captures the way in which part-time workers receive lower pay. The second term, $w(M_t, t)$, is a selection-corrected wage profile taken from French (2000).

The parameters for the idiosyncratic process $\omega_t$, $(\sigma^2_{\eta}, \rho_W)$, shown in Table 3, were estimated in French (1999). The results indicate that the autocorrelation coefficient $\rho_W$ is 0.977; wages are almost a random walk. The estimate of the innovation variance $\sigma^2_{\eta}$ is 0.0141; one standard deviation of an innovation in the wage is 12% of wages. These estimates imply a high degree of long-run earnings uncertainty.

### 4.5 Remaining Calibrations

The remaining calibrations are as follows. When simulating artificial life histories, we use the age-specific interest rate $r_t = r + \varepsilon_t$, to capture the rapid run-up in asset prices that occurred over our sample period. We set the interest rate $r$ equal to 0.04, a standard value. The estimates of $\varepsilon_t$ are described in Appendix F. When finding the decision rules, however, computational concerns led us to set $r_t = r$.\textsuperscript{26} We set the consumption floor $C_{\text{min}}$ equal to $4,000$. This value is almost surely too low; in 1998 the Federal SSI benefit for couples was nearly $9,000 (Committee on Ways and Means, 2000, p. 229). We have chosen to be conservative, because, as discussed below, adding a consumption floor drastically reduces the value of health insurance. Following DeNardi (2000), the object that determines the curvature of the bequest function, $K$, is $500,000$. Spousal income depends upon an age polynomial and the wage. Estimates are available from the authors.

\textsuperscript{26}In applying this differential treatment, we are assuming both that the asset price run-up was unanticipated and that interest rate uncertainty has little effect on saving behavior.
5 Results

5.1 Decision Profiles

We begin with the decision profiles found in the data. Figure 1 shows the 1/3rd and 2/3rd asset quantiles at each age, for both the HRS sample and the model-generated distribution. About one third of the men in the HRS sample live in households with less than $50,000 in assets. Given that some uninsured individuals in the HRS sample had family medical costs in excess of $50,000, there is a non-zero probability that some uninsured low-asset individuals will incur a medical shock that will decimate their financial resources. This indicates that at least some of the HRS sample may value health insurance at well above the dollar cost paid their employers paid for the health insurance. Note, however, that those with assets close to zero will have to pay very little for Medical care if they receive government transfers for medical care. Therefore, it is not clear if low-asset individuals should place a very low valuation on health insurance. The middle third of the asset distribution has assets between $50,000 and $200,000. These individuals also potentially value health insurance at beyond the cost paid by employers. However, it seems unlikely that those at the top of the asset distribution should value health insurance at much beyond cost, as these people should be able to self-insure.

The three panels of Figure 2 show job exit rates by asset quantile and health insurance type. The first panel shows job exit rates for those in the HRS sample whose health insurance is tied to their job. Recall that we would expect Medicare to provide the largest labor market incentives for those workers that have tied health insurance. If these people place a high value on employer-provided health insurance, they should either work until age 65, when they are eligible for Medicare, or they should work until age 63.5 and use COBRA coverage as a bridge to Medicare. Moreover, the size of these incentives should depend on the workers' asset levels. Workers with more assets should be better able to self-insure, while the extremely poor might rely on the insurance implicit in government-provided consumption floors. The first panel of Figure 2, however, shows that the job exit rates differ little by asset level. In fact, high
asset individuals have the highest age-65 job exit rates of any tied health insurance group. This suggests that the value workers place on health insurance does not depend on their asset levels, and is thus roughly equal to the cost paid by employers. These results should be taken with caution, however, because the sample sizes are small. A typical age-asset-quantile cell for those with tied health insurance contains just over 50 individuals.

The middle and bottom panels present job exit rates for workers with retiree coverage and no health insurance, respectively. The most striking feature of these exit rates is how closely they resemble the exit rates for workers with tied coverage. This suggests that access to health insurance is not an important determinant of retirement.

A similar implication appears in Figure 3, which presents labor force participation rates. In comparing the two rates, it is worth noting that job exit rates include only those individuals who were working in the preceding period, while participation rates include everybody. Figure 3 shows that individuals with no employer-provided insurance tend to participate less at every level of wealth. In considering this, it is useful to keep in mind the transitions implied by equation (12): retiring workers in the tied insurance category transition into the none
category. Figure 3 shows that participation rates differ little by wealth. The one exception is the set of workers in the none category, where participation increases with wealth. These wealth-participation relationships may be due to differences in health status, $M_t$—individuals too sick to work cannot accumulate wealth—or differences in wage profiles—the wage profiles of high income/wealth professions are often “backloaded” toward the latter years of the workers’ careers.$^{27}$

5.2 Preference Parameter Estimates

Table 4 presents estimates of the preference parameters, which at this point are still preliminary. Perhaps the most important parameter is $\nu$, the coefficient of relative risk aversion for flow utility. A more familiar measure of risk aversion is the coefficient of relative risk aversion for consumption. This object is not well defined when there are multiple goods. However, assuming that labor supply is fixed and the value of bequests is close to zero, the coefficient of relative risk aversion for consumption can be approximated as 3.29.$^{28}$

$^{27}$We do not account for education and/or profession in our analyses, largely for computational reasons. Among other things, introducing such differences would require us to re-estimate the selection-adjusted profiles described in Section 4.4.

$^{28}$This is measured using the formula $-\frac{(\partial^2 U/\partial C^2)\nu^2}{\partial^2 U/\partial C^2} = -(\gamma(1-\nu) - 1)$. 

27
Figure 2: Job Exit Rates, Data
Figure 3: Labor Force Participation Rates, Data
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>consumption weight</td>
<td>0.596</td>
<td>0.0018</td>
</tr>
<tr>
<td>$\nu$</td>
<td>coefficient of relative risk aversion, utility</td>
<td>6.51</td>
<td>0.055</td>
</tr>
<tr>
<td>$\beta$</td>
<td>time discount factor</td>
<td>1.019</td>
<td>0.003</td>
</tr>
<tr>
<td>$L$</td>
<td>leisure endowment</td>
<td>5627</td>
<td>22.0</td>
</tr>
<tr>
<td>$\phi$</td>
<td>hours of leisure lost, bad health</td>
<td>638</td>
<td>12.4</td>
</tr>
<tr>
<td>$\theta_P$</td>
<td>fixed cost of work, in hours</td>
<td>2294</td>
<td>10.1</td>
</tr>
<tr>
<td>$\theta_B$</td>
<td>bequest weight</td>
<td>0.0572</td>
<td>0.00096</td>
</tr>
</tbody>
</table>

Table 4: structural estimates

5.3 Model Predictions

Figure 1 shows that the model fits both asset quantiles fairly well. The model is able to fit the lower quantile in large part because of the consumption floor; when the consumption floor is lowered from $4,000 to $1, the predicted lower quantile rises dramatically. This is consistent with the results found by Hubbard, Skinner, and Zeldes (1995). Hubbard, Skinner, and Zeldes show that if the government guarantees a minimum consumption level, those with low assets and income will tend not to save, because their consumption will never drop below a certain level, even in the presence of a large negative health cost shock. Put differently, if an individual is at the consumption floor, his savings will be taxed at a marginal rate of 100%. It is therefore not surprising that within the model the consumption floor reduces saving by individuals with low income and assets.

Figure 4 presents simulated job exit rates. Comparing Figures 2 and 4 shows that the model over-predicts job exit rates, especially at age 65. The poor fit is largely an artifact of our MSM estimation procedure; we have chosen to match participation rates rather than job exit rates. The model’s participation rate profiles, shown in Figure 5, match the data much better. Issues of fit aside, the profiles shown in Figures 4 and 5 are consistent with theoretical predictions. In general, workers with higher assets retire earlier. This effect is most pronounced in the participation rates—high asset workers often retire well in advance of Social Security eligibility.
Figure 4: Job Exit Rates, Simulations
Figure 5: Labor Force Participation Rates, Simulations
6 Experiment

With preliminary estimates of all the model’s parameters, we can use the model to assess the relative importance of the normal Social Security retirement age, as opposed to the Medicare eligibility age, in determining retirement behavior. In particular, we shift forward both the normal Social Security retirement age and the Medicare eligibility age from 65 to 67 and observe the increase in simulated work hours. The results of these experiments are summarized in Table 6.

<table>
<thead>
<tr>
<th>Age</th>
<th>Baseline:</th>
<th>Wealth Adj.:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS = 65</td>
<td>SS = 67</td>
</tr>
<tr>
<td>64</td>
<td>626</td>
<td>671</td>
</tr>
<tr>
<td>65</td>
<td>310</td>
<td>585</td>
</tr>
<tr>
<td>66</td>
<td>265</td>
<td>477</td>
</tr>
<tr>
<td>67</td>
<td>241</td>
<td>314</td>
</tr>
<tr>
<td>Total 64-67</td>
<td>1,442</td>
<td>2,046</td>
</tr>
</tbody>
</table>

Table 5: Effects of Changing the Social Security Retirement and Medicare Eligibility Ages

The first column of Table 5 shows average hours worked at ages 64 through 67 under the current retirement and eligibility ages. Under the current rules, the average person works a total of 1,442 hours over this four-year period. The second columns shows that when the normal Social Security retirement age is shifted from 65 to 67 in isolation, the average worker works an additional 605 hours. In addition to changing the rate at which benefits accrue, raising the retirement age effectively eliminates two years of benefits. Under the new regime, workers retiring at any age receive the annual benefit previously given a worker retiring two years earlier. To measure the size of this wealth effect, we raise the retirement age to 67 while increasing annual benefits at every age by 14.8%. The third column shows that under this configuration, the total increases by 283 hours. This suggests that the incentive effects of shifting the retirement age are no bigger than the wealth effects.29

The fourth column of Table 5 shows average hours worked when the Medicare eligibility

---

29In fact, if one considers the total over ages 53-69, the wealth-adjusted retirement age shift actually reduces hours.
age is 67. Raising the eligibility age in isolation increases the total by only 57 hours. The fifth column shows the combined effect of raising both the Social Security retirement and the Medicare eligibility age. The joint effect is an increase of 628 hours, 23 more than that generated by raising the retirement age in isolation. In short, the model predicts that raising the normal Social Security retirement age will have a much larger effect on retirement behavior than shifting the Medicare eligibility age.

In order to understand why Social Security is more important than Medicare, we compute an individual's valuation of health insurance. If individuals do not value health insurance at much beyond the cost paid by employers, the effect of Medicare on retirement behavior will probably be small. We compute the value a worker places on health insurance by computing the increase in assets that would make an uninsured individual as well off as a person with retiree coverage. In other words, we find the value \( \lambda_t = \lambda(A_t, B_t, M_t, AIMEt, \omega_t, \zeta_t, t) \) such that

\[
V_t(A_t, B_t, M_t, AIMEt, \omega_t, \zeta_t, ret) = V_t(A_t + \lambda_t, B_t, M_t, AIMEt, \omega_t, \zeta_t, none).
\]  

Using the health cost processes shown in Table 1, we found \( \lambda(\$146,000,0, good, \$22,000,0, 0, 60) = \$30,100 \). Part of the value of retiree coverage comes from a reduction in average medical expenses, and part of the value comes from a reduction in the volatility of medical expenses. In order to separate the former from the latter, we eliminate health cost uncertainty and recompute \( \lambda_t \), using the same state variables and mean medical expenses as before. Without health cost uncertainty, \( \lambda_t = \$22,600 \).

These two calculations show that \( \frac{\$22,600}{\$30,100} = 75\% \) of the value of health insurance comes from the reduction of average medical expenses, while 25% comes from the reduction in medical cost volatility. This shows why the Medicare has a small effect on retirement behavior. Individuals value health insurance at only slightly more than the cost paid by employers, so that the calculations of Gustman and Steinmeier (1994) are roughly correct.

It is worth noting that employer-provided health insurance has a limited value in large
part because of the government-provided consumption floor. If, for example, the consumption floor is lowered to $1, the value of employer-provided health insurance jumps to $500,000. Our treatment of consumption floors differs markedly from that of Rust and Phelan (1997), who simply impose a penalty when an individual’s implied consumption is negative. Although Rust and Phelan’s estimates do not translate into a consumption floor, they find the penalty to be large, implying a fairly low floor. Since Rust and Phelan assume that consumption equals income net of health costs, it is not clear what their penalties imply for asset accumulation; recall that without a non-trivial consumption floor, our model over-predicts the bottom quantile of assets.

7 Conclusion

Prior to age 65, many individuals receive health insurance only if they continue to work. At age 65, however, Medicare provides health insurance to almost everyone. Therefore, a potentially important work incentive disappears at age 65. If individuals place a high value on health insurance, the provision of Medicare benefits may have a large effect on retirement behavior. To answer this question, we construct and estimate a retirement model that includes health insurance, uncertain medical costs, a savings decision, a non-negativity constraint on assets and a government-provided consumption floor.

Using data from the Health and Retirement Survey, preliminary results suggest that Medicare has a modest effect on retirement behavior. Empirically, we find that job exit rates at age 65 differ little by asset level or health insurance type. This suggests that access to employer-provided health insurance is not an important determinant of retirement. The results from our model reinforce this conclusion. We find that changing the Medicare age in isolation has a relatively small effect on retirement behavior. Moreover, we find that workers value employer-provided health insurance mainly because it reduces average medical expenses, not because it reduces health cost uncertainty. Given that the actuarial value of employer-provided health insurance is fairly modest, this reinforces the conclusion that health
insurance does not provide large retirement incentives.

References


Appendix A: Taxes

Individuals pay federal, state, and payroll taxes on income. We compute federal taxes on income net of state income taxes using the Federal Income Tax tables for “Head of Household” in 1998 with the standard deduction. We also use income taxes for the fairly representative state of Rhode Island (27.5% of the Federal Income Tax level). Payroll taxes are 7.65% up to a maximum of $68,400, and are 1.45% thereafter. Adding up the three taxes generates the following level of post tax income as a function of labor and asset income:

<table>
<thead>
<tr>
<th>Pre-tax Income (Y)</th>
<th>Post-Tax Income</th>
<th>Marginal Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6250</td>
<td>.9235Y</td>
<td>.0765</td>
</tr>
<tr>
<td>6250-40200</td>
<td>5771.9 + .7384(Y-6250)</td>
<td>.2616</td>
</tr>
<tr>
<td>40200-68400</td>
<td>30840.6 + .5881(Y-40200)</td>
<td>.4119</td>
</tr>
<tr>
<td>68400-93950</td>
<td>47425.0 + .6551(Y-68400)</td>
<td>.3449</td>
</tr>
<tr>
<td>93950-148250</td>
<td>64034.4 + .6166(Y-93950)</td>
<td>.3834</td>
</tr>
<tr>
<td>148250-284700</td>
<td>97515.8 + .5640(Y-148250)</td>
<td>.4360</td>
</tr>
<tr>
<td>284700+</td>
<td>174473.6 + .5239(Y-284700)</td>
<td>.4761</td>
</tr>
</tbody>
</table>

Table 6: After Tax Income

Appendix B: Pensions

The fundamental equation behind our calculation of pension benefits is the accumulation equation for pension wealth, \( pw_t \):

\[
pw_{t+1} = \begin{cases} 
(1/s_{t+1})[(1 + r)pw_t + pacc_t - pb_t] & \text{if living at } t + 1 \\
0 & \text{otherwise}
\end{cases}
\]

(23)

where \( pacc_t \) is pension accrual and \( pb_t \) is pension benefits. Two features of this equation bear noting. First, workers cannot bequeath their pensions. It immediately follows that in order to be actuarially fair, surviving workers must receive an above-market return on their pension balances. We achieve this by dividing next period’s pension wealth by the survival probability \( s_{t+1} \) in equation (23). Second, since pension accrual and pension interest are not directly taxed, the appropriate rate of return on pension wealth is the pre-tax one. Pension
benefits, on the other hand, are included in the income used to calculate an individual’s income tax liability.

The rest of this section describes how we compute pension benefits and pension accrual. We calculate pension benefits by assuming that a worker receives no benefits until age 62, at which point his pension is converted into an actuarially fair annuity. To find the annuity, note first that recursively substituting equation (23) and imposing \( pw_{t+1} = 0 \) reveals that

\[
pw_t = \frac{1}{1 + r} \sum_{k=t}^{T} \frac{S(k, t)}{(1 + r)^{k-t}} (pb_k - pac_k),
\]

where \( S(k, t) = (1/s_t) \prod_{j=t}^{k} s_j \) gives the probability of surviving to age \( k \), conditional on having survived to time \( t \). If we assume further that no there is no more pension accrual, so that \( pac_k = 0 \) for \( k = t, t+1, ..., T \), and that pension benefits are constant once they start at age 62, so that \( pb_k = pb_{62} \), \( k > 62 \), this equation reduces to

\[
pw_t = \Gamma_t pb_t, \quad \text{where} \quad \Gamma_t \equiv \frac{1}{1 + r} \sum_{k=t}^{T} \frac{S(k, t)}{(1 + r)^{k-t}} 1\{age_k \geq 62\},
\]

with \( 1\{age_k \geq 62\} \) equal to one if \( age_k \geq 62 \) and equal to zero otherwise. Pension benefits are thus given by

\[
 pb_t = \begin{cases} 
 \Gamma_t^{-1} pw_t & \text{if } t \geq 62 \\
 0 & \text{otherwise} 
\end{cases}.
\]

The pension accrual formula is \( pac_t = pac(W_t H_t, t) \), where

\[
pac(W_t H_t, t) = \alpha_0 \times (\alpha_1 + \alpha_2 W_t H_t + \alpha_3 \max\{0, W_t H_t - 23,000\}) \times \alpha_4(t) \times W_t H_t.
\]

We use a spline function with a kink at \$23,000, \( \alpha_1 + \alpha_2 W_t H_t + \alpha_3 \max\{0, W_t H_t - 23,000\} \), to estimate the dependence of pension accrual on annual labor income. Table 6 of Gustman and Steinmeier (1999) shows that pension accrual rates roughly triple as incomes rise to around
$23,000 per year.\textsuperscript{30} Above this level of income, however, accrual rates are fairly constant. We model the age dependence of accrual rates $\alpha(t)$ using a weighted average of the defined benefit, defined contribution and combined defined benefit and defined contribution profiles shown in Figure 2 of Gustman et al. (1998).\textsuperscript{31} Lastly, we pick the scale parameter $\alpha_0$ so that simulated mean pension wealth is $113,000 in 1998 dollars at age 57, which is meant to coincide with estimates for the male head of household in Table 5 of Gustman and Steinmeier (1999).

To recapitulate, pension wealth follows equation (23), with pension accrual given by equation (27) and pension benefits given by equation (26). Using these equations, it is straightforward to track and record the pension balances of each simulated individual.

But even though it is straightforward to use equation (23) when computing pension wealth in the simulations, it is too computationally burdensome to include pension wealth as a separate state variable when computing the decision rules. Our approach is to impute pension wealth as a function of age and AIME. In particular, we impute a worker’s annual pension benefits as a function of his annual Social Security benefits:

$$\hat{\nu}_t(PIA_t) = \gamma_0(\gamma_1 + \gamma_2 PIA_t + \gamma_3 \max\{0, PIA_t - 6,752\}),$$ \textsuperscript{(28)}

where $PIA_t$ is the Social Security benefit the worker would get if he were drawing benefits at time $t$; as shown in Appendix D below, PIA is a simple monotonic function of AIME. Applying equation (24) yields imputed pension wealth, $\hat{\nu}_t = \Gamma \hat{\nu}_t$. The parameters $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are taken from Gustman and Steinmeier (1999). A spline function is used to estimate $\gamma_1, \gamma_2, \gamma_3$, as Table 6 of Gustman and Steinmeier (1999) shows that the ratio of pension wealth to Social Security wealth rises rapidly with Social Security benefits.\textsuperscript{32} The scale parameter $\gamma_0$ is picked

\textsuperscript{30}Gustman and Steinmeier provide estimates of pension accrual as a function lifetime labor income. We divide lifetime labor income by 30 to get an estimate of average annual labor income.

\textsuperscript{31}We first adjust their pension accrual profile by their assumed rate of wage growth so that pension accrual is measured in rates. We then smooth their pension accrual profile using a 20th order polynomial with dummy variables for age greater than 61, 62, 63, 64 and 65. Predicted accrual rates that are negative are set to zero.

\textsuperscript{32}Gustman and Steinmeier provide estimates of pension wealth and Social Security wealth as functions of lifetime labor income. We convert these measures into annual benefits.
so that mean pension wealth at age 57, as described immediately above, is $113,000.

This imputation process raises some complications. If an individual took his time-$t$ pension wealth as $\widehat{p\omega}_t$, he would assume that his time-$t + 1$ pension wealth (if living) was

$$\widehat{p\omega}_{t+1} = (1/s_{t+1})[(1 + r)p\omega_t + pacc_t - pb_t].$$

This quantity, however, might differ from the pension wealth that would be imputed with $PIA_t$, \( \widehat{p\omega}_{t+1} = \Gamma_{t+1}\hat{p}h_{t+1}. \) To correct for this, we increase non-pension wealth, $A_{t+1}$, by $s_{t+1}(1 - \tau_t)(\widehat{p\omega}_{t+1} - \widehat{p\omega}_t)$. The first term in this expression reflects the fact that while non-pension assets can be bequeathed, pension wealth cannot. The second term, $1 - \tau_t$, reflects the fact that pension wealth is a pre-tax quantity—pension benefits are more or less wholly taxable—while non-pension wealth is post-tax—taxes are levied only on interest income.

A second problem is that while an individual’s Social Security application decision affects his Social Security wealth, it should not affect his pension wealth. The pension imputation procedure we use, however, would imply that it does. We counter this problem by recalculating PIA. In particular, suppose that a decision to accelerate or defer application changes $PIA_t$ to $rem_tPIA_t$. Our approach is to use equation (28) find a value $PIA^*_t$ such that

$$\hat{p}h_t(PIA^*_t) + PIA^*_t = \hat{p}h_t(PIA_t) + rem_tPIA_t,$$

so that the change in the sum of PIA and imputed pension wealth equals just the change in PIA, $(1 - rem_t)PIA_t$.

**Appendix C: Numerical Methods**

Because the model has no closed form solution, the decision rules it generates must be found numerically. We find the decision rules using value function iteration, starting at time $T$ and working backwards to time 1. We find the time-$T$ decisions by maximizing equation (16) at each value of $X_T$, with $V_{T+1} = b(A_{T+1})$. This yields decision rules for time $T$ and the value function $V_T$. We next find the decision rules at time $T - 1$ by solving equation (16) with
$V_T$. Continuing this backwards induction yields decision rules for times $T - 2, T - 3, \ldots, 1$. The value function is directly computed at a finite number of points within a grid, $\{X_i\}_{i=1}^{33}$, we use linear interpolation within the grid and linear extrapolation outside of the grid to evaluate the value function at points that we do not directly compute. Because changes in assets and AIME are likely to cause larger behavioral responses at low levels of assets and AIME, the grid is more finely discretized in this region.

At time $t$, wages, medical expenses and assets at time $t+1$ will be random variables. To capture uncertainty over the persistent components of medical expenses and wages, we convert $\zeta_t$ and $\omega_{t+1}$ into discrete Markov chains, and calculate the conditional expectation of $V_{t+1}$ accordingly. We integrate the value function with respect to the transitory component of medical expenses, $\xi_t$, using 5-node Gauss-Hermite quadrature (see Judd, 1999).

Because of the fixed time cost of work and the discrete benefit application decision, the value function need not be globally concave. This means that we cannot find a worker’s optimal consumption and hours with fast hill climbing algorithms. Our approach is to discretize the consumption and labor supply decision space and to search over this grid. Experimenting with the fineness of the grids suggested that the grids we used produced reasonable approximations. In particular, increasing the number of grid points seemed to have a small effect on the computed decision rules.

We then use the decision rules to generate simulated time series. Given the realized state

\footnote{In practice, the grid consists of: 32 asset states, $A_k \in [-\$45,600, \$900,000]$; 5 wage states, $\omega_i \in [-0.99, 0.99]$; 8 AIME states, $AIME_j \in [\$4,000, \$68,400]$; 3 health cost states, $\zeta_k$, over a normalized \textit{unit variance} interval of $[-1.5, 1.3]$. There are also two application states and two health states. This requires solving the value function at $32 \times 5 \times 8 \times 3 \times 2 \times 2 = 15,360$ different points after age 62, when the individual is eligible to apply for benefits and at 7,680 points before 62.}

\footnote{Using discretization rather than quadrature greatly reduces the number of times one has to interpolate when calculating $E_t(V(X_{t+1}))$.}

\footnote{The consumption grid currently has 100 points. At each value of $X$, the consumption search for time $t$ is centered around the consumption gridpoint that was optimal at time $t+1$. (Recall that we solve the model backwards in time.) For most years, we search over a space that is between 60% and 165% of next period’s consumption rule. For years where there will likely be large changes in the decision rules for a given set of state variables, such as between ages 61 and 62, we increase the search area. If the search yields a maximizing value near the edge of the search grid, the grid is reoriented and the search continued. We begin our search for optimal hours at the level of hours that sets the marginal rate of substitution between consumption and leisure equal to the wage. We then try 6 different hours choices in the neighborhood of the initial hours guess. Because of the fixed cost of work, we also evaluate the value function at $H_t = 0$ where the space of consumption choices is determined by next period’s optimal consumption choice when $H_{t+1} = 0$.}
vector $X_{t0}$, individual $i$’s realized decisions at time 0 are found by evaluating the time-0 decision functions at $X_{t0}$. Using the transition functions given by equations (4) through (13), we combine $X_{t0}$, the time-0 decisions, and the individual $i$’s time-1 shocks to get the time-1 state vector, $X_{t1}$. Continuing this forward induction yields a life cycle history for individual $i$. When $X_{it}$ does not lie exactly on the state grid, we use interpolation or extrapolation to calculate the decision rules. This is true for $\zeta_t$ and $\omega_t$ as well. While these processes are approximated as finite Markov chains when the decision rules are found, the simulated sequences of $\zeta_t$ and $\omega_t$ are generated from continuous processes. This makes the simulated life cycle profiles less sensitive to decision rules at particular values of $\zeta_t$ and $\omega_t$ than when $\zeta_t$ and $\omega_t$ are drawn from Markov chains.

**Appendix D: Computation of AIME**

The Social Security System uses the beneficiary’s 35 highest earnings years when computing benefits. The average earnings over the 35 highest earnings years are called Average Indexed Monthly Earnings, or AIME. We annualize AIME and compute it using the following formula for individuals 30-59.

$$AIME_{t+1} = AIME_t + (W_tH_t)/35.$$  

We assume the individual enters the labor force at age 25. Since AIME is computed using the 35 highest earnings years, AIME increases unambiguously if the individual is younger than 60 and works. If age is 60 or greater AIME can still increase, but only if the individual earns a great deal that year. The high earnings year will replace a low earnings year when computing Social Security benefits.\(^\text{36}\) Therefore, the formula for individuals 60 and older becomes

$$AIME_{t+1} = AIME_t + (W_tH_t - AIME_t)/35. \quad (29)$$

Lastly, AIME is capped. In 1998, the base year for the analysis, the maximum AIME level

\(^{36}\)Unfortunately, we assume that the high earnings year replaces an average earnings year, as described in equation (29).
was $68,400 in 1998 dollars.

AIME is converted into a Primary Insurance Amount (PIA) using the formula

$$PIA_t = \begin{cases} 
0.9 \times AIME_t & \text{if } AIME_t < 5,724 \\
5,151.6 + 0.32 \times AIME_t & \text{if } 5,724 \leq AIME_t < 34,500 \\
14,359.9 + 0.15 \times AIME_t & \text{if } AIME_t \geq 34,500
\end{cases} \quad (30)$$

Social Security benefits $ss_t$ depend both upon the age at which the individual first receives Social Security benefits and the Primary Insurance Amount. For example, pre-Earnings Test benefits for a Social Security beneficiary will be equal to PIA if the individual first receives benefits at age 65. For every year before age 65 the individual first draws benefits, benefits are reduced by 6.7% and for every year (up until age 70) that benefit receipt is delayed, benefits increase by 6%.\textsuperscript{37}

Appendix E: Moment Conditions and the Asymptotic Distribution of Parameter Estimates

We assume that the "true" preference vector $\theta_0$ lies in the interior of the compact set $\Theta \subset \mathbb{R}^7$. Our estimate, $\hat{\theta}$, is the value of $\theta$ that minimizes the (weighted) distance between the estimated life cycle profiles for assets, hours, and participation found in the data and the simulated profiles generated by the model. We match $15T$ moment conditions. They are, for each age $t \in \{1, \ldots, T\}$, two asset quantiles (forming $2T$ moment conditions), labor force participation rates conditional on asset quantile and health insurance type (forming $9T$ moment conditions), labor force participation rates conditional upon health status (forming $2T$ moment conditions), and mean hours worked conditional upon health status (forming $2T$ moment conditions).

Consider first the asset quantiles. As stated in the main text, let $j \in \{1, 2, \ldots, J\}$ index

\textsuperscript{37}AIME can be reduced instead of PIA for individuals who first receive benefits before age 65. For example, if an individual begins drawing benefits at age 62 we can adjust AIME to account for early retirement. We know that adjusted AIME must result in a PIA that is only 80% of what it would have been had the individual first received benefits at age 65. Using equation (30) it is straightforward to compute adjusted AIME. Age at application, then, need not be treated as a state variable.
asset quantiles, where $J$ is the total number of asset quantiles. Assuming that the age-conditional distribution of assets is continuous, the $\pi_j$-th age-conditional quantile of measured assets, $Q_{\pi_j}(A_{it}, t)$, is defined as

$$\Pr \left( A_{it} \leq Q_{\pi_j}(A_{it}, t) | t \right) = \pi_j.$$  

In other words, the fraction of individuals with less than $Q_{\pi_j}$ in assets is $\pi_j$. Therefore, $Q_{\pi_j}(A_{it}, t)$ is the data analog to $g_{\pi_j}(t; \theta_0, \chi_0)$, the model-predicted quantile. Using the indicator function, the definition of $\pi_j$-th conditional quantile can be rewritten as

$$E \left( 1 \{ A_{it} \leq Q_{\pi_j}(A_{it}, t) \} | t \right) = \pi_j. \tag{31}$$

If the model is true then the data quantile in equation (31) can be replaced by the model quantile, so that equation (31) can be rewritten as:

$$E \left( 1 \{ A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0) \} - \pi_j | t \right) = 0, \quad j \in \{1, 2, \ldots, J\}, \ t \in \{1, \ldots, T\}. \tag{32}$$

Equation (32) is merely equation (17) in the text. While equation (32) is a departure from the usual practice of minimizing a sum of weighted absolute errors in quantile estimation, the quantile restrictions just described are part of a larger set of moment conditions. This means that we can no longer estimate $\theta$ by minimizing weighted absolute errors, if only because we are considering multiple quantiles.$^{38}$

We next discuss quantile-conditional means for labor force participation. Let $\mathcal{P}_j(HI, t; \theta_0, \chi_0)$ denote the model’s prediction of labor force participation given asset quantile interval $j$, health insurance type $HI$, and age $t$. If the model is true, $\mathcal{P}_j(HI, t; \theta_0, \chi_0)$ should equal the

$^{38}$A slightly different approach to handling multiple quantiles is the minimum distance framework developed in Epple and Seig (1999). Buchinsky (1998) shows that one could include the first-order conditions from an absolute value minimization problem in the moment set. However, his approach involves finding the gradients of $g_{\pi_j}(t; \theta, \chi)$ at each step of the minimization search.
conditional participation rates found in the data:

$$\mathcal{P}_j(HI, t; \theta_0, \chi_0) = E[P_{it} \mid HI, t, g_{x_{j-1}}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{x_j}(t; \theta_0, \chi_0), \quad (33)$$

with $\pi_0 = 0$ and $\pi_{J+1} = 1$. Equation (33) is equivalent to equation (18) in the text. Using indicator function notation, we can convert the conditional moment equation given by equation (33) into an unconditional one:

$$E[(P_{it} - \mathcal{P}_j(HI, t; \theta_0, \chi_0)) \times 1\{HI_{it} = HI\}$$

$$\times 1\{g_{x_{j-1}}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{x_j}(t; \theta_0, \chi_0)\} \mid t) = 0, \quad (34)$$

for $j \in \{1, 2, ..., J + 1\}$, $HI \in \{none, ret, tied\}$, $t \in \{1, ..., T\}$. Note that $g_{\pi_0}(t) \equiv -\infty$ and $g_{\pi_{J+1}}(t) \equiv \infty$.

Lastly, the moment conditions for labor force participation and hours worked conditional upon health status and age are those described in equations (19) and (20) of the text, converted into unconditional moment equations with indicator functions. Combining all the moment conditions described here is straightforward: we simply stack the moment conditions and estimate jointly.

Next we discuss the distribution of our estimator. Suppose we have a data set of $I$ independent individuals that are each observed for $T$ periods. Let $\tilde{\varphi}(\theta; \chi_0)$ denote the $15T$-element vector of moment conditions that was described in the main text and immediately above. Note that we can extend our results to an unbalanced panel, as we must do in the empirical work, by simply allowing some of the individual’s contributions to $\tilde{\varphi}$ to be “missing”, as in French, (2001). Assuming the model is correctly specified, when $W_T$ is the optimal weighting matrix, the minimized GMM criterion function, $I[1 + \tau]^{-1}\tilde{\varphi}(\theta; \chi_0)'W_TW_T\tilde{\varphi}(\theta; \chi_0)$, is distributed asymptotically as Chi-square with $15T - 7$ degrees of freedom. $\tau$ is the ratio of the number of observations to the number of simulated observations, which tends to zero as the number of simulated observations becomes large. Our estimate of $W_T$ is the inverse of
the 15$T \times 15$ variance-covariance matrix of the data. That is, a typical diagonal element of 
$\hat{W}_T^{-1}$ is the variance estimate $\frac{1}{T} \sum_{i=1}^{T} [1 \{ A_{it} \leq Q_{\pi_j}(A_{it}, t) \} - \pi_j]^2$, while a typical off-diagonal element is a covariance. When estimating preferences, we use sample statistics, so that $Q_{\pi_j}(A_{it}, t)$ is replaced with the sample quantile $\hat{Q}_{\pi_j}(A_{it}, t)$. When computing the chi-square statistic and the standard errors, we use model predictions, so that $Q_{\pi_j}$ is replaced with its simulated counterpart, $g_{\pi_j}(t; \hat{\theta}, \hat{\chi})$. Covariances between asset quantiles and hours and labor force participation are also simple to compute.

Under the regularity conditions stated in Pakes and Pollard (1989) and Duffie and Singleton (1993), the MSM estimator $\hat{\theta}$ is both consistent and asymptotically normally distributed:

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightsquigarrow N(0, V),$$

where $V$ is the variance-covariance matrix of $\hat{\theta}$ that is estimated by

$$\hat{V} = (1 + \tau)(\hat{D}'\hat{W}_T\hat{D})^{-1},$$

$$\hat{D} = \frac{\partial \varphi(\hat{\theta}, \chi_0)}{\partial \theta}.$$ (35)

The gradient in equation (35) is straightforward to compute for hours worked and participation conditional upon age and health status; we merely take numerical derivatives. However, in the case of the asset quantiles and labor force participation, discontinuities make the function $\tilde{\varphi}$ non-differentiable at certain data points. Therefore, our results do not follow from the standard GMM approach, but rather the approach for non-smooth functions described in Pakes and Pollard (1989), Newey and McFadden (1994, section 7) and Powell (1994). We find the asset quantile component of $\hat{D}$ by rewriting equation (32) as

$$F(g_{\pi_j}(t; \theta_0, \chi_0)|t) - \pi_j = 0,$$

where $F(g_{\pi_j}(t; \theta_0, \chi_0)|t)$ is the c.d.f. of time-$t$ assets evaluated at the $\pi_j$-th quantile. Differ-
entiating this equation yields:

\[
D'_{jt} = f(g_{\pi_j}(t; \theta_0, \chi_0) | \mu) \frac{\partial g_{\pi_j}(t; \theta_0, \chi_0)}{\partial \theta}.
\] (36)

In practice we find \( f(g_{\pi_j}(t; \theta_0, \chi_0) | \mu) \), the p.d.f. of time-\( t \) assets evaluated at the \( \pi_j \)-th quantile, with a kernel density estimator.

To find the component of the matrix \( D' \) for the asset-conditional labor force participation rates, it is helpful to write equation (34) as

\[
\Pr(HI_t = HI) \times \int_{g_{\pi_j-1}(t; \theta_0, \chi_0)}^{g_{\pi_j}(t; \theta_0, \chi_0)} [E(P_{it} | A_{it}, HI, t) - \overline{P}_{j}(HI, t; \theta_0, \chi_0)] f(A_{it} | HI, t) dA_{it} = 0,
\]

which implies that

\[
D'_{jt} = \left[ -\Pr(g_{\pi_j-1}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0) | HI, t) \frac{\partial \overline{P}_{j}(HI, t; \theta_0, \chi_0)}{\partial \theta} \\
+ [E(P_{it} | g_{\pi_j}(t; \theta_0, \chi_0), HI, t) - \overline{P}_{j}(HI, t; \theta_0, \chi_0)] f(g_{\pi_j}(t; \theta_0, \chi_0) | HI, t) \frac{\partial g_{\pi_j}(t; \theta_0, \chi_0)}{\partial \theta} \\
- [E(P_{it} | g_{\pi_j-1}(t; \theta_0, \chi_0), HI, t) - \overline{P}_{j}(HI, t; \theta_0, \chi_0)] f(g_{\pi_j-1}(t; \theta_0, \chi_0) | HI, t) \frac{\partial g_{\pi_j-1}(t; \theta_0, \chi_0)}{\partial \theta} \right] \\
\times \Pr(HI_t = HI),
\] (37)

with \( f(g_{\pi_0}(t; \theta_0, \chi_0) | HI, t) \frac{\partial g_{\pi_0}(t; \theta_0, \chi_0)}{\partial \theta} = f(g_{\pi_{j+1}}(t; \theta_0, \chi_0) | HI, t) \frac{\partial g_{\pi_{j+1}}(t; \theta_0, \chi_0)}{\partial \theta} = 0. \)

**Appendix F: Estimating Rates of Return**

This appendix describes how we estimate the rate of return that households faced at different ages in our data. During the sample period of 1992-2000, asset prices grew rapidly. This creates a problem because most of our observations come from the HRS core sample, who were ages 51-61 in 1992, 53-63 in 1994, and so on. Therefore, a 63 year old is more likely to be observed in 1994 than in 1992, whereas a 51 year old is more likely to be observed in 1992 than 1994. In other words, older individuals were typically interviewed in later years, when asset prices were considerably higher. This means that asset levels increase with age in our data not just because of high savings rates, but because of high rates of return. These
rates of return were likely unexpected. In order to control for this, we estimate the ex-post rate of return that individuals faced as they aged.

Let \( r_t \) be the year-specific interest rate faced by a household and let \( \text{age}_{it} \) be the age of individual \( i \) in year \( t \). The average interest rate faced by households headed by an individual of age \( k \) is \( r_k \equiv E[r_t | \text{age}_{it} = k] \). We estimate this object as

\[
\hat{r}_k = \frac{1}{I_k} \sum_{i=1}^{I_k} \sum_{j=1}^{T} r_{tj} \times 1\{\text{age}_{it} = k\},
\]

where \( 1\{\text{age}_{it} = k\} \) is a 0-1 indicator equal to 1 when the age of individual \( i \) in year \( t \) is equal to \( k \), and \( I_k \) is the number of households of age \( k \).

We use two approaches to estimating the year-specific interest rate \( r_t \). In both cases, the idea is to first estimate the historical growth rates in asset prices (we use data over the 1952-2001 period), then take the difference in asset price growth in the 1990s and the historical growth rates. Let \( r_t = E[r_t] + \varepsilon_t \), where \( \varepsilon_t \) represents the the unanticipated component of asset returns, which is white noise. We estimate the anticipated component of asset returns \( E[r_t] \) using the sample mean \( \frac{1}{T} \sum_{t=1}^{T} r_t \). Our goal is to estimate \( \varepsilon_t \).

Our first approach to estimating asset returns to combine estimates of the rate of return on stocks and housing with the shares of household wealth invested in stocks and housing. Let \( A_{lt} \) be the amount of household assets in asset \( l \), \( l \in \{1, ..., L\} \), and let \( r_{lt} \) be the return on that asset. The total return on assets at time \( t \) is

\[
r_t A_{lt} = \sum_{l=1}^{L} A_{lt} r_{lt},
\]

which implies that the average rate of return at time \( t \) is

\[
r_t = \sum_{l=1}^{L} \frac{A_{lt}}{A_{lt}} r_{lt}.
\]
We can then obtain $\varepsilon_t$ by estimating

$$\varepsilon_t = r_t - E[r_t] = \sum_{t=1}^L \frac{A_{1it}}{A_{it}} (r_{1it} - E[r_{1it}]).$$

We estimate $\frac{A_{1it}}{A_{it}}$ using the sample means of shares in different assets.

Most of the unanticipated component of the rate of return over the sample period came from high rates of return in stocks and housing. Assuming that $r_{1it} - E[r_{1it}] = 0$ for all assets other than stocks and housing, we estimate the share of all assets in stocks (19% in our HRS-AHEAD sample) and housing (32% in our sample) and multiply that by the difference between stock returns in year $t, r_{1it}$, less their sample average, $\frac{1}{T} \sum_{t=1}^T r_{1it}$. We estimate the share of stocks in different assets, such as IRAs, using data from the Flow of Funds. The Flow of Funds shows, for example, that in 1995 50% of wealth in Defined Contribution pension plans was in stocks. We estimate rates of return on stocks using data from the Center for Research in Security Prices (CRSP) for the period 1952-2001. Therefore, we multiply .5 by the share of wealth in IRAs to give the share of wealth in stocks held in IRA accounts. We estimate rates of return on housing using data from the Office of Federal Housing Enterprise Oversight for the period 1975-2001. (Earlier data are not available.) These estimated growth rates use data on repeat sales of single family housing that were originated by or subsequently purchased by either Freddie Mac or Fannie Mae.

Our second approach to estimating rates of return is to use aggregate data on savings and asset growth. Aggregate wealth grows according to

$$A_{t+1} = (1 + (r_t(1 - \tau))) A_t + S_t,$$

where $S_t$ is savings between time $t$ and time $t + 1$ and $\tau$ is the average tax rate. Rearranging this equation yields

$$r_t = \frac{A_{t+1} - A_t - S_t}{(1 - \tau) A_t}.$$

We take values of aggregate assets and savings (where savings are defined as personal savings
plus undistributed corporate profits) from the *Flow of Funds* for the period 1952-1991. We assume that \( \tau = 0.2 \).

The second, aggregate data approach is potentially better than the first one because it accounts for the fact that other forms of wealth, such as business wealth, also grew rapidly over the sample period. Moreover, estimating the share of stocks held by households is difficult because it is difficult to infer what share of a household’s IRA wealth is in stocks. Data from the *Flow of Funds* indicate that we are potentially understating the share of total wealth in stocks by over 25% (6.3 percentage points). This would lead us to underestimate the wealth gains from holding stocks over our sample period and thus underestimate \( \varepsilon_t \). The problem with the second approach is that it is very difficult to measure savings. For example, the *Flow of Funds* measure of personal savings is income minus consumption, but income includes rent, dividends and interest. Ideally, the savings measure would be free of rent, dividends and interest. Because firms reduced the share of earnings going to dividends between 1992 and 2000 (leading to higher growth in the value of firms), the data tend to overstate the decline in the savings rate over the sample period. In other words, part of the run-up in assets not explained by savings rates is merely firms buying new equipment instead of paying dividends. This will lead us to overstate the growth in assets not explained by savings and thus \( \varepsilon_t \) over the sample period. Therefore, the two procedures likely provide bounds on \( \varepsilon_t \). Figure 6 shows estimates of \( \varepsilon_t \), by age, over the sample period. The approach using aggregate asset growth results in considerably higher rates of return than does stock returns.
Figure 6: Unanticipated Component of Rate of Return, by Age