THE EFFECTS OF HEALTH, WEALTH, AND WAGES ON LABOR SUPPLY AND RETIREMENT BEHAVIOR

Eric French*
Federal Reserve Bank of Chicago
efrench@frbchi.org

March 16, 2000

Abstract

This paper analyzes the effects of wages and the Social Security System on labor supply over the life cycle. I present a model of labor supply and retirement behavior that includes a savings decision, uncertainty, and a non-negativity constraint on assets. Using data from the Panel Study of Income Dynamics, I estimate life cycle profiles for labor force participation rates, hours worked, and assets. Using the Method of Simulated Moments, I match the estimated profiles to profiles simulated by a dynamic structural model. Estimated parameters produce simulated profiles that match many aspects of the estimated profiles, including the high job exit rates at ages 62 and 65. Simulations suggest that a 20% reduction in Social Security benefits would cause individuals to delay job exit from the labor market in order to develop sufficient financial assets, increasing labor force participation rates from 28% to 35% at age 62.

*Comments welcome. I thank John Kennan, Rody Manuelli, Jonathan Parker, and Jim Walker for detailed comments and encouragement. I also thank Dan Aaronson, Joe Altonji, Peter Arcidiacono, Marco Cagetti, Glen Cain, Kim-San Chung, Meredith Crowley, Nelson Graff, Lars Hansen, Jonathan Hao, John Jones, Yuichi Kitamura, Grigory Kosenok, John Mullany, Derek Neal, Marc Rysman, Karl Scholz, Dan Sullivan, and Bobbi Wolfe. I also thank seminar participants at the University of Wisconsin, Queen’s, Hawaii, Northwestern, the Federal Reserve Bank of Chicago, the Board of Governors, the Bureau of Labor Statistics, Social Security Administration, RAND and the 2000 Boston meetings of the Econometric Society. Tecla Loup and Marita Servais answered many questions about the PSID. Financial support provided by the National Institute on Mental Health. The views of the author do not necessarily reflect those of the Federal Reserve System. Recent versions of the paper can be obtained at http://research.frbchi.org/~efrench/.
1 Introduction

Current demographic changes in the US have increased interest in the question of retirement behavior as policymakers turn increasingly to the question of how to reform the SS system. In order to understand the likely effects of changes in the Social Security System, structural analyses of retirement behavior focus on the dynamic optimization problem that workers face. Previous structural analyses have made very restrictive assumptions about the capital market workers face, however. Many structural retirement models assume individuals are free to save and dissave without limit (Gustman and Steinmeier, 1986; Burtless, 1986). Other retirement models assume that individuals are completely liquidity constrained (Rust and Phelan, 1997; Stock and Wise, 1990). However, given the level of heterogeneity in asset levels of the elderly and the fact that it is illegal to borrow against future Social Security benefits, it seems that neither of these extreme assumptions are correct. Therefore, both of these assumptions will lead to a misspecified model with potentially erroneous predictions.

This paper builds upon the previous approaches. It is the first structural model of labor supply and retirement behavior that includes a savings decision but also has a non-negativity constraint on assets. Moreover, the model accounts for the Social Security rules and the fact that individuals are uncertain about uninsurable future health and wage realizations.

To estimate the model, I use the Method of Simulated Moments (MSM) (Pakes and Pollard (1989), Duffie and Singleton (1993)). The estimation strategy is similar to Gourinchas and Parker (1999). First, I estimate life cycle profiles for assets, hours worked, and labor force participation rates using Panel Study of Income Dynamics (PSID) data. Second, I estimate individual histories of health and wage shocks using PSID data. Third, I simulate individual life cycle profiles for assets, hours worked, and labor force participation rates using the individual histories and the decision rules from the structural model. Fourth, the simulated profiles and the data profiles are aggregated by age and compared to each other. Preference parameters that create simulated profiles that “look like” the profiles from the
data are considered the true preference parameters.

Upon estimating the life cycle profiles for labor supply and labor force participation rates, I find the profiles are hump shaped. Both very young and very old individuals work relatively few hours, with the life cycle profile for hours peaking at around age 35. Labor force participation rates tend to be near 100 percent until age 55, then drop sharply with the largest declines at ages 62 and 65. Using the dynamic structural model, I interpret these profiles as resulting from optimal responses to changes in health, wages, pensions, and the Social Security rules. Labor force participation rates decline 71% between ages 55 and 70. Declining health can explain only a 7% decline in labor force participation rates.

The decline in wages, pension accrual, and the Social Security incentives to retire are a compelling alternative explanation for the decline in labor force participation rates. Like the profile for hours, the profile for wages is hump-shaped, with wages peaking around age 60, and declining sharply after age 65. Likewise, pension accrual declines sharply after age 60. The intertemporal substitution hypothesis suggests that declining compensation should induce exit from the labor force.

Moreover, the Social Security System provides incentives to exit the labor force no earlier than age 62 and no later than 65. Because it is illegal to borrow against Social Security benefits, liquidity constrained individuals must wait until age 62 to draw the benefits that finance their retirement. The Social Security System is actuarially unfair to individuals who wait until after age 65 to draw benefits. Social Security beneficiaries face the Social Security Earnings Test, which potentially leads to a 50% marginal tax rate on labor income. Since the Earnings Test tax is in addition to the usual taxes workers face, Social Security beneficiaries face far higher tax rates than non-beneficiaries. Therefore, the Social Security rules create a strong incentive to exit the labor market by age 65.

The dynamic programming model presented herein captures all of the incentives to exit the labor market. Preference parameters estimated using MSM are reasonable in magnitude. The model captures many features of the data, including the sharp decline in labor force participation rates between ages 55 and 70 and the especially sharp declines at ages 62 and 65.
Simulations suggest that a 20% reduction in Social Security benefits would cause individuals to delay job exit from the labor market by two months in order to develop sufficient financial assets, increasing labor force participation rates from 28% to 35% at age 62.

The rest of the paper is as follows. Section 2 develops a model of optimal lifetime decision-making. Since there is no closed-form solution to the model, I solve the model using numerical methods, described in Section 3. Section 4 describes the estimation scheme: the Method of Simulated Moments. Section 5 describes the data. Section 6 presents parameter estimates. Section 7 describes the policy experiments. Section 8 concludes.

2 The Model

This section describes the model of lifetime decision-making. Individuals choose consumption, work hours (including the labor force participation decision) and whether or not to apply for Social Security benefits. They are allowed to save, although assets must be non-negative. When making these decisions, they are faced with several forms of uncertainty: survival uncertainty, health uncertainty, and wage uncertainty.

Individuals who are heads of households are assumed to maximize expected lifetime utility from $t = 1$ to $T + 1$, where $t$ is measured in years. The probability of being alive at time $t$ is denoted $S_t$ and the probability of being alive at time $t$ conditional on being alive at time $t - 1$ is $s_t$. Therefore, $S_t = \prod_{r=1}^{t} s_r$ where $s_r = (1 - \text{prob}(\text{dead}_r | \text{alive}_{r-1}))$. Since time $T + 1$ is the terminal period, $s_{T+1} = 0$. $\beta$ is the time discount factor. Individuals value bequests of assets, $A_t$, according to a bequest function $b(A_t)$. The within-period utility function $U_t = U(C_t, H_t, M_t)$ depends on consumption, $C_t$, hours worked, $H_t$, and health (or medical) status, $M_t$, which alters preferences for hours worked. Individuals also choose whether to apply for Social Security benefits, $B_t \in \{0, 1\}$ which is a 0-1 indicator equal to one if the individual has applied for benefits.

The individual’s maximization problem can be written as:
\[
\max_{C_t, H_t, B_t} \mathbb{E}_t \left[ \sum_{j=t}^{T+1} \beta^j S_{j-1} \left( s_j U(C_j, H_j, M_j) + (1 - s_j) b(A_j) \right) \right]
\]  

subject to a mortality determination equation (4), a health determination equation (5), a wage determination equation (6), and an asset accumulation equation (8). Note that if the individual does not value bequests both the survivor function and the time discount factor enter the maximization problem in the same way. If the consumer perceives a large probability of death in the near future, the consumer will be more impatient and will save less for the future.

The within-period utility function is of the form

\[
U(C_t, H_t, M_t) = \left( C_t^{1-\gamma_c} \frac{1}{1-\gamma_c} + \phi_M \frac{(L - H_t - \theta P_t)^{1-\gamma_h}}{1-\gamma_h} \right).
\]

where the per period time endowment is \( L \) and

\[
\phi_M = \begin{cases} 
\phi_g & \text{if } M_t = \text{good} \\
\phi_b & \text{if } M_t = \text{bad}
\end{cases}
\]

The parameters \( \gamma_c \) and \( \gamma_h \) are the coefficients of relative risk aversion for consumption and hours, respectively. 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, \( \theta P_t \), is measured in hours worked per year.\(^1\) Retirement is assumed to be a form of the participation decision. Workers

\(^1\)Annual hours of work is clustered around both 2000 hours and 0 hours of work, a regularity in the data that standard utility functions have a difficult time replicating. Fixed costs of work are a common way of explaining this regularity in the data (Cogan (1981)). Fixed costs of work generate a reservation wage for a given marginal utility of wealth. Below the reservation wage, hours worked is zero. Slightly above the reservation wage, hours worked may be large. Individual level labor supply is highly responsive around this reservation wage level, although wage increases above the reservation wage result in a smaller labor supply response.
can reenter the labor force.

The bequest function is of the form

\[ b(A_t) = \theta_B \frac{(A_t + K)^{1-\gamma_c}}{1 - \gamma_c}. \]  

(3)

where \( K \) determines the curvature of the bequest function. If \( K = 0 \) there is infinite disutility of leaving non-positive bequests. If \( K > 0 \) the utility of a zero bequest is finite.

Given the objective function, individuals face several constraints. Mortality rates depend upon age\(^2\) and previous health status:

\[ s_{t+1} = s(M_t, \text{age}_{t+1}). \]  

(4)

Health status at time \( t, M_t \) can take two values, good and bad. Next year’s health status \( \text{prob}(M_{t+1}|M_t, \text{age}_{t+1}) \) depends on current health status and age. Health status follows a two-state transition matrix at each age with a typical element\(^3\)

\[ \pi_{\text{good, bad}, t+1} = \text{prob}(M_{t+1} = \text{good}|M_t = \text{bad, age}_{t+1}). \]  

(5)

Individuals can move from good health to bad health and from bad health to good health.

The logarithm of wages\(^4\) at time \( t, W_t \), is a function of age and health status, \( b(M_t, \text{age}_t) \), plus an autoregressive\(^5\) component of wages \( AR_t \):

\[ \text{The notation } \text{age}_t \text{ is redundant as both } \text{age}_t \text{ and } t \text{ are measured in years, but I make the distinction for the sake of clarity.} \]

\[ \text{I ignore the possibility that wealth may affect health, as the Grossman (1972) model implies.} \]

\[ \text{By “wage,” I am referring to the observed wage of labor market participants as well as the potential wage of non-participants. Given this definition of wage, another interpretation for the wage would be “productivity.”} \]

\[ \text{This follows most studies of wage dynamics (Farber and Gibbons, 1996, French, 1999) which find that wages follow a highly persistent } AR(1) \text{ process.} \]
\[ \ln W_t = h(M_t, age_t) + AR_t, \]  \hfill (6)

The autoregressive component of wages has a correlation coefficient \( \rho \) and a normally distributed innovation \( \eta_t \):

\[ AR_t = \rho AR_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma^2). \]  \hfill (7)

By assumption, at time \( t-1 \) the worker knows the autoregressive component of wages \( AR_{t-1} \) but only knows the distribution of the innovation in the wage \( \eta_t \) in the subsequent time period.

The final constraint is the asset accumulation equation:

\[ A_{t+1} = A_t + Y(rA_t + W_t H_t, \tau) + p_t + ss_t - C_t, \quad A_{t+1} \geq 0, \]  \hfill (8)

where \( Y(rA_t + W_t H_t, \tau) \) is the level of post tax income, \( r \) is the interest rate, \( \tau \) is the tax structure (described in Appendix A), \( p_t \) is pension accrual, and \( ss_t \) is Social Security benefits.

The pension accrual formula \( p_t \) 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 are in Appendix B.

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 \( B_t = 1 \). Application for Social Security benefits is only a decision if the individual is older than 62 and has not
yet applied for benefits. Upon application for benefits the individual receives benefits until death, i.e. \( B_{t+1} = 1 \) if \( B_t = 1 \). Only if the individual is 62 or older but has not yet applied for benefits is the benefit application decision is a decision variable. Once the individual has applied for Social Security benefits, benefits depend on Average Indexed Monthly Earnings, or \( AIME_t \), which is roughly the 35 highest earnings years in the labor market. Section (5.2) describes specifics of how \( AIME_t \), age of application for benefits, and labor income affects \( ss_t \). Appendix B describes computation of \( AIME_t \).

Also, individuals face a non-negativity constraint on assets. Because it is illegal to borrow against Social Security benefits, individuals with low assets potentially must wait until age 62 to finance exit from the labor market.

3 Numerical Methods

This section outlines the methods for computing the decision rules. Specifically, it outlines the methods for computing the value function, the methods for integrating the value function with respect to uncertainty over wages, and the method to find the optimal consumption and hours decisions.

Optimal decisions depend on the state variables, \( X_t = (A_t, W_t, B_t, M_t, AIME_t) \), preferences denoted \( \theta = (\gamma_c, \gamma_h, \theta_P, \theta_B, \phi_g, \phi_b, \beta) \), and beliefs denoted \( \chi = (r, v^2, \rho, h(M_t, age_{t+1}), \{prob(M_{t+1}|M_t, age_t)\}_{t=1}^T, \{S_t\}_{t=1}^T, Y(\cdot, \cdot), \{p_t\}_{t=1}^T, \{ss_t\}_{t=1}^T) \). The value function is the solution to

\[
V_t(X_t) = \max_{C_t, H_t, B_t} \left\{ \left( \frac{C_t^{1-\gamma_c}}{1-\gamma_c} + \phi_M \frac{(L - H_t - \theta_P P_t)^{1-\gamma_h}}{1-\gamma_h} \right) + \beta s_{t+1} \sum_M \int V_{t+1}(X_{t+1})dF(W_{t+1}|M_{t+1}, W_t, t)prob(M_{t+1}|M_t, t) + \beta(1 - s_{t+1})b(A_{t+1}) \right\}, \tag{9}
\]

where \( dF(\cdot|\cdot, \cdot, \cdot) \) is the conditional cdf of next period’s wages. The individual is uncertain of future wage and health shocks. The values of \( C_t, H_t \) and \( B_t \) that solve (9) are considered the
optimal consumption and hours decisions.

Although the consumption, hours, and participation and benefit application rules have no closed-form solutions, the rules fully characterize the decisions of the individual. The solution to the worker’s problem then consists of a 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 which solve the value function (9). The labor force participation rule \( \{ P_t(X_t, \theta, \chi) \}_{1 \leq t \leq T} \) is equal to one if \( \{ H_t(X_t, \theta, \chi) \}_{1 \leq t \leq T} = 0 \) and equals zero otherwise. Using these decision rules and the asset accumulation equation it is also possible to solve for next period’s asset level \( \{ A_{t+1}(X_t, \theta, \chi) \}_{1 \leq t \leq T} \).

The decision rules can then be solved for recursively, starting at time \( T \) and working backwards to time \( 1 \). I compute the value function using value function iteration. At time \( T \), consumption and hours decisions will be made by maximizing equation (9), where \( V_{T+1} = b(A_{T+1}) \). Consumption and hours decisions are next solved for time \( T-1, T-2, T-3, ..., 1 \) by backwards induction. Using this technique the individual decision rules at time \( t \) can be found as functions of only the state variables at time \( t \).

Since there is no closed form solution to the problem, the value function is evaluated at a finite number of points within a grid, \( \{ X_i \}_{i=1}^I \). Because variation in assets, AIME and wages is likely to cause larger behavioral responses at low levels of assets, AIME and wages, the grid is more finely discretize at low levels of assets, AIME and wages. Since the value function is computed at a finite number of points, I use linear interpolation within the grid and extrapolation outside of the grid to evaluate the value function points that were not directly computed.

I integrate the value function with respect to the innovation in the wage using Gauss-Hermite quadrature. Although assets at time \( t+1 \) will be known at time \( t \), wages at time \( t+1 \) will be a random variable. In practice, I use quadrature of order 5 (Judd, 1999).

I also discretize the consumption and labor supply decisions and use a grid search tech-

---

6In practice, I chose 32 asset states, 14 wage states. The grid for assets and wages is \( A_i \in [80, 700,000] \), \( W_j \in [83, 86] \). There are two application states, two health states, and three possible benefit levels. This requires solving the value function at \( 32 \times 14 \times 2 \times 2 \times 3 = 5,376 \) different points after age 62 when the individual is eligible to apply for benefits and \( 2,688 \) points when younger than 62.
nique to find the optimal consumption and hours rules. Because the fixed cost of work and the benefit application decision mean that the value function need not be globally concave, I cannot use relatively fast hill climbing algorithms. I experimented with the fineness of the grids. The grids described herein seemed to produce reasonable approximations.\footnote{Currently, there are 90 possible values for consumption. I use next period's optimal consumption rule as an initial guess for this period's optimal consumption rule at each value of $X$. For most years, I search over a space that is between 70\% and 150\% 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, I increase the search area to 30\%-300\% of next period's optimal decision rule. To find the optimal hours decision, I use the marginal rate of substitution between consumption and leisure. There is a difficulty in that the marginal rate of transformation between consumption and leisure is not the wage. Instead, taxes, pensions, and the effect of current work hours on Social Security benefits distort the relationship. Therefore, I make an initial guess by setting the marginal rate of substitution equal to the wage. I then try 10 different hours choices in the neighborhood of the initial hours guess. Because the fixed cost of work may cause large discontinuous changes in optimal hours worked (from zero hours worked and a large number of hours worked), I 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 $H_{t+1} = 0$.}

4 Estimation

This section describes the method of simulated moments (MSM) estimation strategy. The goal is to estimate the preferences $\theta$ given beliefs $\chi$. Because it would be too computationally burdensome to estimate both preferences and beliefs contemporaneously, I use a two-step strategy. In the first step I estimate some belief parameters and calibrate others. I assume rational expectations, meaning that individuals know their own state variables $X_t$ at time $t$, the Markov process that determines their state variables, which is parameterized by $\chi$, and optimize accordingly. The second step uses the numerical methods described in the previous section and the belief parameters to simulate life cycle profiles for simulated individuals. Preferences that generate simulated profiles that match profiles estimated from data are considered to be true preferences. The next subsection describes the MSM technique. The following subsection describes construction of the sample profiles.

4.1 Estimation of Preferences: The Method of Simulated Moments

The MSM estimation strategy matches means of the assets, hours of work and the participation decision to means of the same variables in a simulated sample. The “matching” of
moments is done using standard GMM techniques. Because of problems with measurement error, I do not match high order moments.\textsuperscript{8} Using means, however, averages out measurement error, as shown below.

The MSM procedure allows for heterogeneity in the state variables, \( X_{it} \), but the requirement of computational simplicity does not allow for heterogeneity in preferences \( \theta \) or beliefs \( \chi \). I assume that individual \( i \) responds only to \((X_{it}, \theta, \chi)\). Wages and health status may differ across individuals, which means that their expected future wages and health statuses will be different, but the stochastic processes that determine health and wages are the same. Differences in wages and health status imply that there may be differences in decisions.

The objective is to find a vector of preferences \( \theta_i \in \Theta \) that simulates profiles that “look like” (as measured by a GMM criterion function) the profiles from the data. I assume \( \Theta \subset \mathbb{R}^7 \) where \( \Theta \) is a compact set. I postulate the following model for the data-generating process:

\[
A_{it} = A_t(X_{it-1}, \theta, \chi) + \epsilon_{it},
\]

\[
\ln H_{it} = \ln H_t(X_{it}, \theta, \chi) + \epsilon_{iHt} \quad \text{if} \quad P_{it} > 0,
\]

\[
P_{it} = P_t(X_{it}, \theta, \chi)
\]

where \( A_{it}, \ln H_{it} \) and \( P_{it} \) represent individual \( i \)’s measured assets, log of hours worked, and participation decisions at time \( t \), and \( \epsilon_{it}, \epsilon_{iHt} \) represent measurement error in consumption and hours. I assume zero mean measurement error in assets and hours, \( E[\epsilon_{it} | A_t(X_{it-1}, \theta, \chi)] = 0 \) and \( E[\epsilon_{iHt} | H_t(X_{it}, \theta, \chi)] = 0 \). Computation of \( A_t(X_{it-1}, \theta, \chi), \ln H_t(X_{it}, \theta, \chi) \) and \( P_t(X_{it}, \theta, \chi) \) is described in the Numerical Methods section of the paper.

The difference between the data (e.g., \( A_{it} \)) and the predictions of the model (e.g., \( A_t(X_{it-1}, \theta, \chi) \)) arises only from measurement error if the state variables \( X_{it-1} \) are the same. If there is no

\textsuperscript{8}See Altonji (1986), Abowd and Card (1989) and French (1998) for attempts to overcome the measurement error problems that plague high frequency analyses of labor supply.
measurement error in age or health status then it is possible to generate moment conditions
for participation and hours worked conditional upon health status, resulting in the following
5T moment conditions:

\[
E[A_{it} - A_t(X_{it-1}, \theta, \chi)|M_{it}, t] = E[A_{it}] - \int A_t(X, \theta, \chi)dF_{t-1}(X) = 0, t \in \{1, ..., T\},
\]

\[
E[\ln H_{it} - \ln H_t(X_{it}, \theta, \chi)|M_{it}, t] =
E[\ln H_{it}] - \int \ln H_t(X, \theta, \chi)dF_{it}(X|M) = 0, t \in \{1, ..., T\}, M \in \{\text{good, bad}\},
\]

\[
E[P_{it} - P_t(X_{it}, \theta, \chi)|M_{it}, t] =
E[P_{it}] - \int P_t(X, \theta, \chi)dF_{it}(X|M) = 0, t \in \{1, ..., T\}, M \in \{\text{good, bad}\},
\]

where \(F_t(X)\) is the cdf of the state variables at time \(t\) and \(F_{Mt}(X|M)\) is the cdf of the state
variables at time \(t\) given health status \(M\). At the true preference parameters and at the state
variables \(X\) which have the same distribution within the data as within the simulated sample,
conditional on age and, in the case of hours and participation, health status, the expectation
of the difference between the data sample and the simulated sample is zero.

In summary, the MSM procedure I use can be described as follows. First, estimate life
cycle profiles from the data. Second, using the same data used to estimate the profiles,
generate matrices for random health and wage shocks as well as an initial distribution for
health, wages and assets. The matrices include shocks for 5,000 simulated individuals, and
over their entire lives. Third, pick an arbitrary \(\theta_i \in \Theta\) and compute the decision rules given \(\theta_i\)
and the numerical methods described in the previous section. The fourth step is to simulate
profiles for the decision variables. Using the decision rules and the health and wage shocks,
profiles of decision variables are simulated for each simulated individual given the health and
wage shocks over the life cycle. The simulated data are aggregated by age (and in the case of hours and participation, health status) in the same way the true data are aggregated by age. Sixth, moment conditions are computed. That is, the distance between the simulated and true profiles is then computed. Finally, a new value of \( \theta_i \in \Theta \) is picked and the whole process is repeated.\(^9\) The value of \( \theta_i \) that minimizes the distance between the true data and the simulated data described in equations (13)-(15) is considered the true value of \( \theta, \hat{\theta} \). I leave the distribution of the parameter estimates, the weighting matrix and the overidentification tests to Appendix E.

4.2 Estimation of Profiles

This section describes the life cycle profiles for assets, hours, and participation rates to be fed into equations (13) - (15) as well as the life cycle wage profile. When constructing profiles, I am concerned about the presence of individual-specific effects and family size effects as well as age and health effects. Let \( Z_{it} \in \{ \ln W_{it}, \{ \ln H_{it} \}, \{ P_{it} \}, \{ A_{it} \} \} \) denote an observation of wages or a decision variable for individual \( i \) at age \( t \) who heads a family of family size \( famsize \). I consider the following model:

\[
Z_{it} = f_i + \Pi_{age} c_{it} \times \text{prob}(M_{it} = \text{good} | M_{it}^a) + \Pi_{age} c_{it} \times \text{prob}(M_{it} = \text{bad} | M_{it}^a) + \Pi_{famsize} + u_{it} \tag{16}
\]

where \( \text{prob}(M_{it} = \text{bad} | M_{it}^a) \) is the probability that health is bad given a noisy health measure \( M_{it}^a \) and \( \text{prob}(M_{it} = \text{good} | M_{it}^a) = 1 - \text{prob}(M_{it} = \text{bad} | M_{it}^a) \). I describe construction of \( \text{prob}(M_{it} = \text{bad} | M_{it}^a) \) in Appendix D. If \( M_{it}^a \) were perfectly measured, then \( \text{prob}(M_{it} = \text{bad} | M_{it}^a) \) would collapse to the standard dummy variable (i.e. \( \text{prob}(M_{it} = \text{bad} | M_{it}^a) = 1(M_{it}^a = \text{bad}) \)) were \( M_{it}^a \) measured without error). I estimate equation (16) using fixed-effects to control for the individual-specific effect, \( f_i \). I use a full set of age dummy variables when estimating the hours, participation, and asset profiles, although the wage profile I feed into the MSM.

\(^9\)I use simplex methods. Specifically, I use the amoeba, written by Bo Honore and Elaterini Kyriasidou to search over \( \Theta \).
algorithm is smoothed using a polynomial in age. I use a full set of dummy variables for family size $famsize$.

I use the age effects and health effects from equation (16) to generate the profiles. I set family size equal to three, and use the mean individual-specific effect for individuals who were born in 1940, who are age 50, and have the average level of health for 50 year olds when generating profiles. This point is described in greater detail in Appendix E. I use the fourth order age polynomial to generate the profiles for wages. For the decision variable profiles, I use the age dummy variables. For the asset profiles I assume $\Pi_y = \Pi_h$, as I do not condition on health status when generating the asset profile.

Unfortunately, the fixed effects estimator does not overcome selection problems in the wage equation as estimates from the wage equation are for observed wages for workers and does not include the potential wages of non-workers. Because fixed-effects estimates growth rates for wages and not just levels, composition bias problems (i.e. the question of whether high wage or low wage individuals drop out of the labor market) are addressed if the growth rates for workers and non-workers are the same. If growth rates for wages for workers and non-workers is the same, it is sufficient to know only the growth rates for workers to infer the growth rates for non-workers. However, if individuals leave the market because of a sudden wage drop, such as from job loss, I will not be able to include the new potential wage of those individuals. This problem will bias wage growth upwards.

5 Data and Calibrations

5.1 Data

In order to estimate the model proposed in the previous three sections I use the Panel Study of Income Dynamics (PSID) for the years 1968-1997. I drop the SEO subsample. The PSID tracks households, not individuals. Therefore, I use labor supply variables for the male

---

$^{10}$When creating profiles with the polynomials, I estimate the polynomial using data on individuals five years younger and 10 years older than my sample of interest. This overcomes some of the endpoint problems associated with polynomial smoothing.
head of household but use household-level asset data.

When estimating the hours worked and labor force participation rate profiles, I use individuals born between 1922 and 1940, resulting in 18,690 person-year observations for labor force participation rates and 15,766 person-year observations for hours worked. For the asset profile, I use individuals born between 1902 and 1965 to increase sample size, resulting in 8,265 person year observations. For the wage and health profiles, I use the full sample.

I estimate the asset profile using 1984, 1989, and 1994 PSID wealth surveys. Because I do not wish my estimate of assets to be affected by the extremely wealthy, many of whom inherit their wealth, I exclude observations with over $1,000,000 in assets. Households in which entering family member brought assets into the household or an exiting family member took assets out of the household are dropped. The PSID asset measure is fairly comprehensive. It includes real estate, the value of a farm or business, vehicles, stocks, mutual funds, IRAs, Keoghs, liquid assets, bonds, other assets and investment trusts less mortgages and other debts. It does not include pension and Social Security wealth. Although the simulated asset measure does not include Social Security wealth, which is modeled formally, it does include pension wealth. Therefore, the PSID measure of wealth is matched to simulated wealth net of the value simulated pension wealth where the simulated wealth measure is described in appendix B.

Wages are computed as annual earnings divided by hours and are dropped if wages are less than $3 per hour or greater than $100 per hour. Hours are counted as zero if measured hours are below 500 hours worked per year.

The PSID has only one measure of health that is asked during all years of the panel. It is the self-reported response to “Do you have any physical or nervous condition that limits the type of work or the amount of work that you can do?” A criticism of self-reported health measures is that respondents often report “bad health” in order to justify being out of the labor force. This will lead me to overestimate the effect of health upon work hours. Alternatively, the coarse discretization of health status into good and bad when true health status is likely a continuous variable potentially causes measurement error, biasing the effect
of health status on different variables to zero effect.

The PSID has poor information on mortality statistics. Therefore, I combine PSID data with mortality statistics from the National Center for Health Statistics (NCHS).\textsuperscript{11} These statistics use the entire US population as their sample.

5.2 Social Security

There are three major incentives provided by the Social Security System.\textsuperscript{12} All three incentives tend to induce exit from the labor market when old. First, because Social Security benefits depend upon Average Indexed Monthly Earnings (AIME), which is the 35 highest earnings years, increased income when young results in increased Social Security benefits when old. This generates work incentives when young that tend to offset the payroll tax disincentives for work for workers during their first 35 years in the labor market. Computation of AIME is described in Appendix C.

Second, the age at which the individual applies for Social Security benefits affects the level of benefits when receiving benefits. Individuals are ineligible for Social Security benefits before age 62. Between the ages of 62 and 65 the Social Security System is roughly actuarially fair but is unfair after age 65. This causes people to apply for benefits by age 65. Social Security benefits are a progressive function of AIME and the year the individual starts drawing benefits. For every year before 65 the individual applies for benefits, benefits are reduced by 6.7%. This is roughly actuarially fair. For every year after age 65 that benefit application is delayed, benefits rise by 3% up until age 70. This is actuarially unfair.

Third, the Social Security Earnings Test taxes labor income for Social Security benefi-

\textsuperscript{11}I compute mortality rates given last year’s health status using Bayes’ rule:

\[ prob(\text{death}_t | M_{t-1} = \text{good}) = \frac{\text{prob}(M_{t-1} = \text{good} | \text{death}_t)}{\text{prob}(M_{t-1})} \times \text{prob}(\text{death}_t) \]  

I compute \( \text{prob}(M_{t-1} = \text{good} | \text{death}_t) \) and \( \text{prob}(M_{t-1}) \) using PSID data, and \( \text{prob}(\text{death}_t) \) from the NCHS data. When using PSID data, the estimate of \( \text{prob}(\text{death}_t) \) is about 25% lower than when using NCHS data, indicating that the PSID underestimates mortality rates by 25%.

\textsuperscript{12}I use tax and benefit formulas from the Social Security Handbook Annual Statistical Supplement for the year 1987 for several reasons. Most importantly, 1987 is relatively close to the middle year of the data. Second, there were significant simplifications to the tax code enacted in 1986 that simplifies the dynamic programming problem. Lastly, benefit formulas have not become significantly more or less generous after 1987.
ciaries at a very high rate. This causes a reduction in post tax wages for Social Security beneficiaries. The incentive to draw benefits by age 65 in combination with the Social Security Earnings Test for Social Security beneficiaries is a major disincentive for work after age 65. Individuals younger than age 70 face the Social Security Earnings Test. Above a “test” threshold level of $6,000, benefits are taxed at 50% until all benefits have been taxed away. If current benefits are taxed away, future benefits are recomputed. If a year’s worth of benefits are taxed away between 62 and 65, benefits in the future will be recomputed upwards by 6.7%. If a year’s worth of benefits are taxed away between 65 and 70, benefits in the future will be recomputed upwards by 3%.

The asset accumulation equation (8) captures all these incentives.

A common misconception is that the recomputation formulas fully replace benefits lost through the Earnings Test. Although this is roughly true between ages 62 and 65, a loss of one month’s benefits results in only a small upward revision in future benefits after age 65. Moreover, the Earnings Test tax on benefits is in addition to Federal, state, and payroll taxes on income so labor income is taxed twice. Therefore, the marginal tax rate an individual faces will potentially be greater than 50%. For example, an individual facing a 25% marginal tax rate for labor income and a 50% marginal tax on benefits faces a 75% effective marginal tax rate.

5.3 Remaining Calibrations

In order to estimate preferences, I calibrate some components of the belief vector $\chi$. These calibrations include information on the stochastic component of wages $(v^2, \rho)$, the interest rate $r$, and the initial level of assets. The parameters from the wage equation $(v^2, \rho)$, shown in

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>Estimate</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v^2$</td>
<td>variance of the innovation in wages</td>
<td>.0141</td>
<td>.0014</td>
</tr>
<tr>
<td>$\rho$</td>
<td>autoregressive coefficient of wages</td>
<td>.977</td>
<td>.017</td>
</tr>
</tbody>
</table>

Table 1: The Variance and Persistence of Innovations to the Wage

Table 1, have been estimated in French (1999). The results indicate that $\rho = .977$; wages are
almost a random walk. The estimate of \( v^2 \) is .0141; one standard deviation of an innovation in the wage is 12% of wages. These estimates imply that long run forecast errors may be large.

The remaining calibrations are as follows. I set the interest rate at \( r = .05 \). Following Ghez and Becker (1975), the per year time endowment is \( L = 5096 \) hours. Following DeNardi (2000), the object that determines the curvature of the bequest function, \( K \), is $500,000.

6 Results

The estimated inputs into the MSM algorithm can be divided into data on beliefs and data on decisions. The belief parameters \( \chi \) include growth rates for wages conditional on health status, health transition matrices, and mortality probabilities. The decision profiles include hours worked per year by workers, assets, and Labor Force Participation Rates. In order to distinguish between the alternative theories of wages versus health in explaining the decline of hours near the end of the life-cycle, profiles for hours and labor force participation rates are shown for individuals in both good and bad health.

6.1 Belief Profiles

The belief parameters include parameters from the wage equation, the health transition matrices, and the survivor probabilities. In this section, I describe these profiles. Although I use smoothed versions of the profiles when estimating preferences, I display unsmoothed profiles to show how precisely the profiles are measured. Overall, the profiles are precisely measured, as displayed by their smooth appearance. On average, profiles for healthy individuals are smoother than for unhealthy individuals. This is because there are more observations on healthy individuals than on unhealthy individuals.
Figure 1: Life Cycle Profiles for Belief Variables
Using the methodology from Section 4.2, the top left panel of Figure 1 displays wage profiles for males by age and health status. Most striking is the hump shape of the wage profiles for both health groups, with wages peaking near age 55. Also striking is the small effect of health on wages. Fixed-effect estimates show a smaller role for health than OLS. There are three alternative explanations for the difference between OLS estimates and fixed-effect estimates. First, it may be that some other factor (e.g. childhood poverty) causes both poor health and low wages. Second, the Grossman model (1972) predicts that individuals who have higher lifetime wages invest more in health human capital when young. Therefore, the Grossman model implies that high wages cause good health, and not vice versa as most interpretations of an OLS regression of wages on health assume. The third explanation for the small estimated effect of health upon wages could be related to a selection problem. It may be only the individuals who get lucky in the labor market who remain in the labor market after a bad health shock. Given that one standard deviation of the innovation to wages is about 12% of wages, it is likely that true wages for unhealthy individuals are somewhere between 0% and 12% lower than what the top left panel suggests.

The two right hand side variables of Figure 1 shows how health dynamics change over the life-cycle. Until age 55, very few individuals experience a change from good health to bad health. This begins to change after age 55, with individuals becoming more and more likely to move from good health to bad health. Note, however, there is no rapid shift in population health that takes place only between ages 55 and 70, the ages at which labor force participation declines most rapidly. Instead, much of the decline in population health takes place after age 70.

Lastly, the lower left panel shows mortality rates over the life cycle. Unsurprisingly, individuals in bad health have higher mortality rates than individuals in good health.

6.2 Decision Profiles

Given the beliefs in the previous section, I will now describe the profiles for the decision variables. These profiles will identify the preference parameters in the utility function. I
make two fundamental identifying assumptions. First, the belief parameters are exogenous. Second, preferences depend only upon health and family size. Preferences change with age, but only as a result of changes in health and family size. Therefore, age can be thought of as an “exclusion restriction” which causes changes in beliefs but not preferences.

The top two panels of Figure 2 show the life-cycle profiles for hours worked and participation rates for men in good and bad health. The effect of health on hours worked is sizeable, but can only explain a small amount of the variation in work hours over the life cycle. Hours worked begins to decline rapidly after age 55. This is true even when conditioning on health status, so it appears that health status alone must have a small causal role in the decline in the number of hours worked near retirement.

Health appears to affect labor force participation rates more. However, the fraction of all individuals at age 55 who report bad health is 20% and this rises to only 37% by age 70. Therefore, the change in labor force participation rates attributable to changes in health between ages 55 and 70 is

$$\sum_{t=56}^{70} \Delta P_t = \sum_{t=56}^{70} \left( \frac{\Delta P}{\Delta M} \right)_t \Delta M_t$$

(18)

where $\Delta$ is the standard first difference operator, $\left( \frac{\Delta P}{\Delta M} \right)_t$ is the estimated effect of health on participation (estimated by the vertical difference between the upper and lower profiles in the middle panel in Figure 2) at age $t$, and $\Delta M_t$ is the change in population health status between age $t - 1$ and $t$ (estimated using the bottom right panel of Figure 1). This technique suggests that declining health between ages 55 and 70 can explain a 7% drop in labor force participation rates. Thus, of the drop in labor force participation rates from 87% to 13% between ages 55 and 70, only 10% can be attributed to declining health. Moreover, the ages at which hours and labor force participation rates decline most rapidly coincides with those ages at which wages begin and at which there are large Social Security work disincentives. Therefore, it seems that wages and Social Security potentially play a strong role in determining the age of retirement.
Figure 2: Life Cycle Profiles for Decision Variables
Finally, the bottom panel of Figure 2 shows assets over the life cycle. Note that young people do save. A certainty life cycle model with typical parameters predicts that people dissave when young since wage levels are very low when young. Therefore, the life cycle asset profile is evidence against the standard certainty-equivalent life cycle model. However, it is consistent with a model in which young people save in order to generate a buffer stock of assets for insurance against bad wage shocks when old (Cagetti, 1999).

6.3 Initial Distributions

The joint distribution between wages and assets is assumed log-normal at age 30. Average assets at age 30 are equal to $42,100 (which is the average level of assets for households aged 28-32) and are highly correlated with wages. The variance of log assets\(^{13}\) at age 30 is .95 and the variance of log wages is .23.\(^{14}\)

6.4 Preference Parameter Estimates

Table 2 presents estimates of the parameters in the utility function for males, ages 30 - 95. Because of lack of data on older individuals, I assume that individuals do not work after age 80 and match moments only up to age 70. Table 2 presents structural estimates of the parameters within the utility function.

The coefficient \(\gamma_h\), the coefficient of relative risk aversion for leisure, is equal to the inverse of the intertemporal elasticity of leisure if the individual is working. I estimate the intertemporal elasticity to be .6, larger than most estimates that use cohort data (Ghez and Becker, 1975, and Browning et al., 1985, estimate the intertemporal elasticity to be around .3).\(^{15}\) Moreover, the estimate of the intertemporal elasticity of substitution is larger for

\(^{13}\) Measured using assets greater than $5,000.

\(^{14}\) This initial variance in wages will partially control for educational status. Since wages are highly persistent, those who have higher wages in the first period will, on average, have higher wages in the final period than individuals who had low wages in the initial period. Given this, the model captures the fact that college graduates have higher wages than non-graduates.

\(^{15}\) Many of these studies estimate the intertemporal elasticity of labor supply, not leisure. Ghez and Becker estimate both and find that the intertemporal elasticity of labor supply is slightly greater than the intertemp-
groups where many people are not working, such as people in their 50s and 60s, because of the fixed cost of work. At the "reservation" level of work hours, the intertemporal elasticity of substitution is infinity. The simulated elasticity of average hours worked given an anticipated transitory change in the wage is .5 at age 40 and 1.5 at age 60.\textsuperscript{16}

<table>
<thead>
<tr>
<th>Parameter and Definition</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_c$ coefficient of relative risk aversion, consumption</td>
<td>1.42</td>
<td>0.002</td>
</tr>
<tr>
<td>$\gamma_h$ coefficient of relative risk aversion, leisure</td>
<td>1.74</td>
<td>0.003</td>
</tr>
<tr>
<td>$\beta$ time discount factor</td>
<td>0.989</td>
<td>0.001</td>
</tr>
<tr>
<td>$\phi_g$ leisure preference parameter, good health</td>
<td>0.0157</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\phi_b$ leisure preference parameter, bad health</td>
<td>0.219</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\theta_F$ fixed cost of work, in hours</td>
<td>1259</td>
<td>16</td>
</tr>
<tr>
<td>$\theta_B$ bequest weight</td>
<td>38.9</td>
<td>1.3</td>
</tr>
</tbody>
</table>

$\chi^2$ statistic: 676 (193 degrees of freedom)

Table 2: STRUCTURAL ESTIMATES

My estimate of the intertemporal elasticity of substitution is identified primarily by data on older workers. Figure 2 shows hours worked and labor force participation rates declining rapidly after age 60, even after the effect of health on hours has been addressed. This decline in hours coincides closely with the decline in wages and the Social Security incentives to retire. Therefore, evidence from older individuals indicates that labor supply is responsive to changes in economic incentives and thus the intertemporal elasticity of substitution is large. Many of the studies that use the life cycle profiles for male work hours to estimate the intertemporal elasticity of substitution (Ghez and Becker, 1975, and Browning et al., 1985) obtain identification from a problematic source: the covariation of work hours and wages of continuously employed young workers. Young workers work many hours although on average their wage is low. This indicates that the intertemporal elasticity of substitution is small within a certainty-equivalent environment, as hours change very little but wages change a lot over the life cycle. However, uncertainty means that younger workers may work many hours in order to develop enough assets to buffer themselves against the possibility of low wages

\textsuperscript{16}This calculation was made by changing the wage by 20% for all workers of at a given age, then computing the difference in total hours worked at that age.

poral elasticity of leisure. This shows my large estimate is not driven by using the elasticity of leisure instead of hours.
when old. Omission of uncertainty will potentially bias the estimated intertemporal elasticity of substitution downwards.

The coefficient $\gamma_s$, the coefficient of relative risk aversion (or the inverse of the intertemporal elasticity) for consumption, is similar to previous estimates that rely on different methodologies (see Auerbach and Kotlikoff (1987) for a review of the estimates). Identification of this parameter is similar to Cagetti (1999) who estimates a buffer stock model of consumption over the life cycle using asset data. Within this framework, a small estimate of the coefficient of relative risk aversion means that individuals save little given their level of assets and their level of uncertainty. If they were more risk averse (i.e. if they had a larger $\gamma_s$), they would save more in order to buffer themselves against the risk of bad income shocks in the future. I also obtain identification from labor supply data, as precautionary motives can explain why wages co-vary little with hours when young but a great deal when old. If individuals were more risk averse, they would work more hours when young. Working more hours when young allows risk averse individuals to generate a buffer stock of assets.

My estimate of the time discount factor, $\beta$, is larger than most estimates of $\beta$. There are three reasons for such a large estimate. First, most studies do not include mortality risk. In my model, individuals discount the future not by the discount rate $\beta$, but by the discount factor multiplied by the survivor function $s_t$. Since the survivor function is necessarily less than one, omitting mortality risk will bias $\beta$ downwards. Second, the post-tax rate of return is smaller than the pre-tax rate of return. Since the return to saving is lower than a model without taxes, individuals must be more patient in order to explain why people save. Third, the life cycle profile of hours shows that young individuals work many hours even though their wage, on average, is low. This indicates that young people buy relatively little leisure when young, even though the price of leisure (their wage) is low. Between ages 35 and 60, people buy more leisure as they age even though their price of leisure (or wage) increases. A large estimate of $\beta$ is required to reconcile these facts.

---

17This is exactly correct when individuals do not value bequests.
Figure 3: Simulated Profiles Versus True Profiles
Figure 3 displays both the data and the simulations. It also displays 95% confidence intervals (i.e. plus or minus two standard deviations). Simulations appear generally consistent with the data, although the simulated profiles frequently lie outside of the confidence intervals. A $\chi^2$ overidentification test rejects the model because the simulated profiles lie outside of the confidence intervals. There are some differences between the simulations and data that are worthy of mention. The model overpredicts asset levels. This problem results from the high estimate of the patience parameter $\beta$. Since simulated individuals in the model are extremely patient, they wish to consume more when old than when young. This results in a high savings rate when young, causing high asset levels when old. Moreover, the model overpredicts job exit rates at age 62. Simulated individuals with high asset levels appear to be less responsive to the age 62 incentives than simulated individuals with low asset levels. Therefore, it appears that at estimated parameter values the model displays strong liquidity effects on labor force participation rates—stronger than what is seen in the data.

There are two reasons for the small standard errors in Table 2. First, the GMM criterion function formulae for standard errors relies on the assumption that the GMM criterion function is quadratic near the minimum of the function. This is true in the case of a linear model, but may be a poor approximation in the case of a non-linear model. This results in standard errors being underestimated in a non-linear model.\(^{18}\) Second, small standard errors may result from the assumption that the first stage estimates of $\chi$ are measured with no error. Gourinchas and Parker (1999) suggest a method that allows one to incorporate the variance of the first stage parameter estimates into the standard errors to the second stage estimates. Note, however, that the profiles for both the decisions and beliefs are precisely estimated as shown by their smooth appearance. This arises from the extremely large sample size used in the analysis. Moreover, using data on older workers and labor force participation rates greatly increases the variation in the data. Therefore, it is unsurprising that standard

---

\(^{18}\)To address this problem, I tried an alternative technique to obtain a measure of the precision of the estimates. I adjusted $\gamma_h$ upwards by 5% to 1.49 and re-estimated the other parameters. A $\chi^2$ test of the unrestricted model versus the model with $\gamma_h$ set at 1.49 resulted in only a narrow rejection of the restricted model (the difference in the $\chi^2$ statistics was 4.1, with a critical 5% value of 3.8). This tends to show that while the standard errors are being underestimated, the model is sharply identified.
errors are smaller than other analyses using PSID data (e.g., MaCurdy (1981)).

6.5 Robustness Checks

Perhaps the most striking prediction of the model is the rapid decline in labor force participation rates at the exact age of 62. There are several reasons why these drops could happen within the model: the rapid decline in pension accrual after age 62, actuarial unfairness of the Social Security System, and liquidity constraints. The model of pension accrual allows for discontinuous jump in pension accrual at ages 61, 62, 63, 64, and 65. However, even after forcing pension accrual to be smooth, the age 62 downturn in labor force participation rates is still large (as is the age 65 downturn). Discontinuities in pension accrual explain less than 20% of the decline in labor force participation rates at age 62 within the model.

Whether the Social Security System is actuarially unfair depends heavily upon the assumed rate of interest. Given a 5% pre-tax interest rate (resulting in a 3-4% post tax rate of return), Social Security benefit accrual is negative (and thus actuarially unfair) at age 62. However, even after using a 1% rate of interest, making Social Security benefit accrual positive between ages 62 and 65, the decline in labor force participation rates at age 62 still exists. Actuarial unfairness explains very little of the decline.

One final robustness check tests the sensitivity of the estimates to the exogenous wage assumption. Gustman and Steinmeier (1986) suggest that part-time workers are paid less than full-time workers. I re-estimated the wage profile assuming that part-time (1000 hours per year) workers are paid 25% less per hour than full-time (2000 hours per year) workers, resulting in a wage profile displaying almost no wage declines after age 60. Estimated preferences were fairly similar to the estimates in Table 2 where wages do not depend on hours worked, except that $\beta$ was larger. Since wages did not decline after age 60, patience becomes the explanation for why individuals exit the labor force. The simulated profiles did a worse job of matching the estimated profiles.
7 Experiments

A question of policy concern is “how does Social Security generosity affect labor supply of the elderly?” I conduct a partial equilibrium\textsuperscript{10} experiment in which I reduce Social Security benefits by 20\%. Figure 4 displays the simulated profiles from the experiment.

Figure 4 displays results from simulations under both the current policy environment and the environment in which benefits have been reduced by 20\%. Reducing Social Security benefits increases hours worked after age 50, although most of the effect is after age 62.

Reducing Social Security benefits will cause a decrease in lifetime wealth. This should result in individuals working more hours throughout their lives, as individuals consume less leisure with the loss of wealth. Figure 4 shows that individuals will possibly work more hours and increase assets to offset reduced benefits when old. At age 62, simulated labor force participation rates are 28\% with current benefit levels and 35\% with reduced benefit levels. Total years in the labor force between ages 30 and 70 rise from 30.75 to 30.92 with the reduction in benefits. The reduction in benefits leads to greater wealth accumulation as individuals develop sufficient financial resources to finance retirement. These effects, while non-trivial, should be interpreted as being relatively small. Simulations suggest a more extreme experiment where the entire Social Security System is eliminated results in an increase of years in the labor force between ages 30 and 70 from 30.75 to 32.44 years. In other words, total elimination of the Social Security System would delay retirement less than two years.

\textsuperscript{10} These simulations do not account for the fact that the United States government would have to adjust tax rates and benefit levels to pay for the changes in Social Security expenditures. Instead, these simulations capture the effect of reduced Social Security generosity in isolation. This approach can be criticized since lower benefits would result in lower government expenditures. Lower tax revenues would be needed to finance these expenditures. However, the focus of this paper is to better understand the effect of changes in government policy on individuals' decisions. Also, the Social Security system is a pay-as-you go system where the taxes of younger generations pay for the benefits of older generations. Total taxes paid by a generation are not directly linked to total benefits received by the same generation. Therefore, imposing a balanced budget constraint seems inappropriate.
Figure 4: The Effect of Reduced Benefits
8 Conclusion

This paper presents estimates from a dynamic structural model of life-cycle labor supply and retirement behavior. Preferences are estimated using the Method of Simulated Moments (MSM). The analysis is novel in several respects. Most importantly, it is the first structural life cycle model that includes a retirement decision, a savings decision, and a non-negativity constraint on assets. The model also includes uncertainty, detailed modeling of the Social Security System, and a decision to re-enter the labor force. Lastly, it considers the decisions of men aged 30-90, a much longer time span than most analyses.

I find that declining health can explain a 7% decline in labor force participation rates between ages 55 and 70, although labor force participation rates fall by 71% during the same age span. Moreover, I find that wages decline sharply after age 60. This, coupled with declines in pension accrual and the actuarial unfairness of the Social Security System after age 65, provides a strong incentive for retirement. Using these incentives, the dynamic programming model described herein explains the decline in labor force participation well. It also does a reasonable job of explaining the life cycle profiles of work hours by workers and assets. Moreover, the estimated parameters are reasonable in magnitude. For example, the labor supply elasticity to an anticipated transitory change in the wage .5 at age 40 and 1.5 at age 60. This higher elasticity at older ages results from the labor force participation margin.

The value of such a model lies in its ability to predict how labor supply and retirement patterns of individuals might change in response to changes in the Social Security rules. Simulations suggest that a 20% drop in Social Security benefits results in an increase in labor supply throughout the life cycle in order to replace lost Social Security benefits. The primary source of this increase is in delaying age of exit from the labor force, as workers increase their years in the labor force between the ages of 30 and 70 from 30.75 to 30.92 years. Simulated labor force participation rates rise from 28% to 35% at age 62. This would result in increased asset accumulation by workers between the ages of 62 and 70 as they develop the necessary assets to finance retirement.
Because this model is so rich in detail, it could be usefully applied to a number of questions. For example, it would be interesting to simulate the effect of increasing the normal Social Security retirement age from 65 to 67. Another interesting line of research would be to explain the decline in male labor force participation rates in the 1970s. The relative importance of changes in pensions, Disability Insurance, Medicare, Social Security, and the wage structure in determining the reduction in age of retirement in the 1970s remains unresolved. Current retirement models include only a subset of these institutional constraints. The approach presented herein can be adapted to include all of them.

References


Appendix A: Taxes

Individuals pay federal, state, and payroll taxes on income. I compute federal taxes on income net of state income taxes using the Federal Income Tax tables for “Head of Household” in 1987 with the standard deduction. I also use income taxes for the fairly representative state of Rhode Island (22.96% of the Federal Income Tax level). Payroll taxes are 7.15% up to a maximum of $43,000. 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>0-4440</td>
<td>0.9285Y</td>
<td>0.072</td>
</tr>
<tr>
<td>4440-6940</td>
<td>4123 + 0.796(Y-4440)</td>
<td>0.204</td>
</tr>
<tr>
<td>6940-27440</td>
<td>6113 + 0.749(Y-6940)</td>
<td>0.251</td>
</tr>
<tr>
<td>27440-42440</td>
<td>21468 + 0.6021(Y-27440)</td>
<td>0.398</td>
</tr>
<tr>
<td>42440-43800</td>
<td>30500 + 0.5262(Y-42440)</td>
<td>0.474</td>
</tr>
<tr>
<td>43800-84440</td>
<td>31216 + 0.5977(Y-43800)</td>
<td>0.403</td>
</tr>
<tr>
<td>84440+</td>
<td>55506 + 0.5605(Y-84440)</td>
<td>0.440</td>
</tr>
</tbody>
</table>

Table 3: After Tax Income

Appendix B: The Pension Accrual Formula

The pension accrual formula is $p_t = p(W_t H_t, age_t)$:

$$p(W_t H_t, age_t) = \alpha_0 \times (\alpha_1 + \alpha_2 W_t H_t + \alpha_3 \max(0, W_t H_t - 23,000)) \times \alpha_4 (age_t) \times W_t H_t.$$  \hspace{1cm} (19)

I pick the scale parameter $\alpha_0$ so that mean pension wealth$^{20}$ is $101,207$ in 1987 dollars at age 60, which is meant to coincide with estimates in Table 6 of Gustman and Steinmeier (1999). I 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)$^{21}$ shows that pension accrual rates roughly triple between individuals with extremely small incomes and individuals with incomes around $23,000$. Above this level, however, accrual rates are fairly constant. Lastly, I model the age dependence of accrual rates $\alpha_4 (age_t)$ using a weighted average of the defined benefit, defined contribution and combined defined benefit and defined contribution profiles in Figure 2 of Gustman et al. (1998)$^{22}$.

Appendix C: Computation of AIME

$^{20}$ I measure pension wealth using the formula

$$pensionwealth_t = \sum_{k=1}^{i} (1 + r(1 - \tau))^k p(W_k H_k, age_k)$$  \hspace{1cm} (20)

where $\tau$ is the marginal tax rate at an individual's face.

$^{21}$ They provide estimates of pension accrual as a function lifetime labor income. I divide lifetime labor income by 30 to get an estimate of average annual labor income.

$^{22}$ I adjust their pension accrual profile by their assumed rate of wage growth so that pension accrual is measured in rates 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.
The Social Security System uses the beneficiaries 35 highest earnings years when computing benefits. The average earnings over the 35 highest earnings years are called Average Indexed Monthly Earnings, or AIME. I annualize AIME and compute it using the following formula for individuals 30-59.

\[ AIME_{t+1} = AIME_t + \frac{(W_t H_t)}{35}. \]  

(21)

I 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.\(^{23}\) Therefore, the formula for individuals 60 and older becomes

\[ AIME_{t+1} = AIME_t + \frac{(W_t H_t - AIME_t)}{35}. \]  

(22)

I assume AIME does not increase upon receipt of benefits.\(^ {24}\) Lastly, AIME is capped. In 1987, the base year for the analysis, the maximum AIME level was $43,800 in 1987 dollars.

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

\[ PIA_t = \begin{cases} 
.9 \times AIME_t & \text{if } AIME_t < 3,720 \\
3,348 + .32 \times AIME_t & \text{if } 3,720 \leq AIME_t < 22,392 \\
9,695 + .15 \times AIME_t & \text{if } AIME_t \geq 22,392 
\end{cases} \]  

(23)

Social Security benefits \( s_t \) depend both upon the age at which the individual first receives

\(^{23}\)Unfortunately, I assume that the high earnings year replaces an average earnings year, as described in equation (22).

\(^{24}\)This assumption is not correct. However, since work hours of Social Security beneficiaries is usually low, this assumption is fairly innocuous. It has the advantage in that it allows me to omit age at first receipt of benefits as a state variable. 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 81% of what it would have been had the individual first received benefits at age 65. Using equation (23) it is straightforward to compute adjusted AIME. Age at application, then, need not be treated as a state variable.
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 3%. If the individual first began drawing benefits at age 64, the pre-Earnings Test benefit level for this person would be $0.933 \times PIA$, since his PIA would be reduced by 6.7%. This point, as well as the Earnings Test and benefit recomputation formulas, have been discussed in the text.

Appendix D: Estimation of Misclassification Error in Health Status

In this appendix I outline the approach to estimating the extent of misclassification error in my health measure. I use the estimates of the extent of misclassification error when estimating life cycle profiles of hours worked, labor force participation rates, and wages. Failure to control for misclassification error when estimating the life cycle profiles in equation (16) will cause an errors-in-variables problem that tends to bias the estimated effect of health variables towards zero effect. However, it is possible to obtain consistent estimates of the effect of health on hours worked, labor force participation rates, and wages using fixed effects if the extent of measurement error is known. Misclassification error can arise from two sources.

First, misclassification error can arise from misreports on the part of the respondent or transcription error on the part of the enumerator. This means that measured health may be bad when true health is good, or vice versa. Note that this measurement error is not zero mean when conditioning on measured health status since a report of good health can only imply a correct classification of good health or that true health is bad.

Second, the health measure I use corresponds to a point in time whereas the hours variable corresponds to hours over the previous calendar year. Therefore, an individual who reports good health at the time of the interview may not be in good health over the entire year. The procedure described in this section overcomes both the problem of misreporting and the problem of timing.

To address the first problem of misreports, I estimate the misclassification propensities
by using two health measures. Conceptually, the approach is the same as the approach of Bound (1991) or Stern (1989), but I account for the fact that measurement error may not be zero mean.

The primary health measure that I use is the response to “do you have any physical or nervous condition that limits the amount or type of work that you can do?” Throughout I will refer to this measure of health as the PSID measure. I wish to estimate the misclassification rate of this measure. To do so, I use a supplement to the PSID called the 1990 Self Administered Questionnaire (SAQ). In 1990, individuals age 50 and older were asked a large number of questions about their health. The SAQ was not part of the normal battery of questions in the PSID. Instead, the questionnaire was done by mail. The questionnaire was sent slightly after the normal interview. The response rate to this survey was 74.4%. I took six questions from this survey, all designed to measure whether a physical or nervous condition limits the amount or type of work the respondent can do.\(^{25}\) If the respondent reported yes to any of the health limitation measures, they are considered to have reported bad health. Throughout I will refer to this measure of health as the SAQ measure. The PSID health measure and the SAQ health measure are very similar in the question asked. As a result, the probability of a response of bad health in both measures is similar. In the estimation sample of 693 respondents, 35% reported bad health using the usual PSID measure and 33% reported bad health using the SAQ measure.

The fact that the two measures of health were taken at slightly different times and the fact that the two questions were designed to measure the same object motivates two key assumptions. First, I assume that misclassification errors in usual PSID health measure and the SAQ health measure are independent of one another. The fact that the two measures were taken at slightly different times should mitigate problems misclassification error in the two measures being correlated because an individual was having a “bad day” and reported bad health on both measures. Second, I assume that the two health measures have the

\(^{25}\)Specifically, the questions are v63-v68 in the SAQ. For example, v63 asks “During the past month, have you cut down on the amount of time you could spend on work or other activities as a result of your physical health?”
same misreporting properties. I describe what I mean by similar misreporting properties below. This assumption seems reasonable given that both the PSID measure and the SAQ measure ask similar questions. Furthermore, they have the same probability of a “bad health” response. Lastly, I make the assumption that (conditional upon age) the two health measures are unbiased. That is, at each age the expected number of healthy people who errantly report unhealthy when healthy is the same as the expected number of people who errantly report healthy when unhealthy.

Let true health for individual $i$ at time $t$ be $M_{it}$. Denote the PSID measure of health as $M_{it}^{PSID}$ and the SAQ measure as $M_{it}^{SAQ}$. Let $M_{it} = 1$ if the individual’s true health is bad and let $M_{it} = 0$ if their true health is good. Let $M_{it}^{PSID} = 1$ if the individual’s measured health in the PSID is bad and let $M_{it}^{PSID} = 0$ if their measured health in the PSID is good. Likewise, $M_{it}^{SAQ} = 1$ and $M_{it}^{SAQ} = 0$ denotes measured bad health and good health in the SAQ, respectively.

Define $\gamma_t$ as the age-specific probability that the respondent reports good health when in bad health and define $\delta_t$ as the age-specific probability that the respondent reports bad health when in good health. Therefore,

$$prob(M_{it}^{PSID} = 1|M_{it} = 1) = prob(M_{it}^{SAQ} = 1|M_{it} = 1) = 1 - \gamma_t$$

$$prob(M_{it}^{PSID} = 0|M_{it} = 0) = prob(M_{it}^{SAQ} = 0|M_{it} = 0) = 1 - \delta_t$$

Further, define $\kappa$ as the age invariant rate of total misreports, i.e., $\kappa = \delta_t + \gamma_t$.

Using the Law of Total Probability

$$prob(M_{it}^{PSID} = 1) = prob(M_{it}^{PSID} = 1|M_{it} = 1)prob(M_{it} = 1) + prob(M_{it}^{PSID} = 1|M_{it} = 0)prob(M_{it} = 0) = (1 - \gamma_t)prob(M_{it} = 1) + \delta_t(1 - prob(M_{it} = 1)),$$
and the unbiasedness assumption, \( \text{prob}(M_{it} = 1) = \text{prob}(M_{it}^{PSID}) \), equation (26) can be rewritten as

\[
\delta_t = \kappa \text{prob}(M_{it}^{PSID} = 1),
\]

\[
\gamma_t = \kappa - \delta_t = \kappa \text{prob}(M_{it}^{PSID} = 0).
\]

Since I assume the SAQ measure of health has the same misclassification properties as the PSID measure, equation (27) holds for the SAQ measure of health as well. Intuitively, what equation (27) says is that in order for \( M_{it}^{PSID} \) to be an unbiased estimator of \( M_{it} \) (i.e., for \( \text{prob}(M_{it} = 1) = \text{prob}(M_{it}^{PSID} = 1) \)), the fraction of people in a particular age group who report unhealthy when healthy is increasing in the fraction of the age group who reports unhealthy. To be unbiased there must be a healthy person who reports unhealthy for each unhealthy person who reports healthy. As the fraction of the age group who are unhealthy \( \text{prob}(M_{it} = 1) \) increases, there must be a growing fraction of all people who are healthy who report unhealthy. This implies that older workers, for whom the probability of bad health \( \text{prob}(M_{it}) \) is higher, are more prone to reporting bad health when in good health.

Using the Law of Total Probability again, the joint probability that both measures of health are bad is

\[
\text{prob}(M_{it}^{PSID} = 1, M_{it}^{SAQ} = 1) = \text{prob}(M_{it}^{PSID} = 1, M_{it}^{SAQ} = 1|M_{it} = 1)\text{prob}(M_{it} = 1) + \text{prob}(M_{it}^{PSID} = 1, M_{it}^{SAQ} = 1|M_{it} = 0)\text{prob}(M_{it} = 0)
\]

\[
= \text{prob}(M_{it}^{PSID} = 1|M_{it}^{SAQ} = 1, M_{it} = 1)\text{prob}(M_{it}^{SAQ} = 1|M_{it} = 1)\text{prob}(M_{it} = 1) + \text{prob}(M_{it}^{PSID} = 1|M_{it}^{SAQ} = 1, M_{it} = 0)\text{prob}(M_{it}^{SAQ} = 1|M_{it} = 0)\text{prob}(M_{it} = 0)
\]

Equation (29) can be simplified greatly. Since by assumption the misclassifications in the two measures of health are uncorrelated, \( \text{prob}(M_{it}^{PSID} = 1|M_{it}^{SAQ} = 1, M_{it} = 1) = \text{prob}(M_{it}^{PSID} = 1|M_{it}^{SAQ} = 1) \)
\[ \text{prob}(M_{it}^{P\text{SID}} = 1, M_{it}^{SAQ} = 1) = (1 - \gamma_t)^2 \text{prob}(M_{it} = 1) + \delta_t^2 \text{prob}(M_{it} = 0). \] (30)

Finally, using equation (27), we may rewrite equation (30) as

\[
\text{prob}(M_{it}^{P\text{SID}} = 1, M_{it}^{SAQ} = 1) = \kappa^2 \text{prob}(M_{it}^{P\text{SID}} = 1|M_{it} = 1)^2 \text{prob}(M_{it}^{P\text{SID}} = 1|M_{it} = 0) \\
+ (1 - \kappa \text{prob}(M_{it}^{P\text{SID}} = 1|M_{it} = 0))^2 \text{prob}(M_{it}^{P\text{SID}} = 1|M_{it} = 1)
\] (31)

which may be further simplified to

\[
\text{prob}(M_{it}^{P\text{SID}} = 1, M_{it}^{SAQ} = 1) = \text{prob}(M_{it}^{P\text{SID}} = 1) + \kappa(\kappa - 2)\text{prob}(M_{it}^{P\text{SID}} = 1)\text{prob}(M_{it}^{P\text{SID}} = 0)
\] (32)

Note that there is only one unknown in equation (32): the variable \( \kappa \). This is the variable we wish to estimate. By replacing the equality sign in equation (32) with a minus sign we have the object we wish to minimize:

\[
\min_{\kappa} \left\{ \text{prob}(M_{it}^{P\text{SID}} = 1, M_{it}^{SAQ} = 1) - \right. \\
\left. \left( \text{prob}(M_{it}^{P\text{SID}} = 1) + \kappa(\kappa - 2)\text{prob}(M_{it}^{P\text{SID}} = 1)\text{prob}(M_{it}^{P\text{SID}} = 0) \right) \right\}
\] (33)

Applying similar reasoning results in a set of moment conditions using the joint probability that an individual is in good health using both health measures:

\[
\min_{\kappa} \left\{ \text{prob}(M_{it}^{P\text{SID}} = 0, M_{it}^{SAQ} = 0) - \right. \\
\left. \left( \text{prob}(M_{it}^{P\text{SID}} = 0) + \kappa(\kappa - 2)\text{prob}(M_{it}^{P\text{SID}} = 1)\text{prob}(M_{it}^{P\text{SID}} = 0) \right) \right\}
\] (34)

Equations (33) and (34) also hold for when \( M_{it}^{SAQ} \) is used on the right hand side instead of \( M_{it}^{P\text{SID}} \). Since I have data on individuals from ages 50-79, this results in 30 moment conditions each for equations (33) and (34) for both the SAQ and the standard PSID measures. This
results in $30 \times 4 = 120$ moment conditions.

My estimate of $\kappa$ is .29 with a standard error of .02. This means that for each age $\delta_t + \gamma_t = .29$, indicating that about 15% of all health reports are incorrect.

Using equations (27) and (34), the value of $\kappa$ can be used to predict the probability that someone is in good health given their report.

\[
prob(M_{it} = 1|M_{it}^{PSID} = 1) = 1 - \gamma_t = 1 - \kappa prob(M_{it} = 0) \tag{35}
\]
\[
prob(M_{it} = 0|M_{it}^{PSID} = 0) = 1 - \delta_t = 1 - \kappa prob(M_{it} = 1). \tag{36}
\]

To address the second measurement problem, the timing issue, I use a simple interpolation scheme. Most PSID interviews are in April or May. The PSID health measure refers to current health whereas the hours measure refers to hours worked in the previous calendar year. This means that an individual who is in bad health at the time of the interview will not necessarily be in bad health for the full calendar year. The following procedure addresses the timing problem.

Addressing this problem requires additional notation. Define $t$ as the time period from Jan. 1 to Dec. 31, which is the standard notion of time throughout the paper. Let $t^*$ denote the month of the interview, $t^* \in \{1, \ldots, 12\}$. Then $M_{it,t^*}$ refers to health status during month $t^*$ of year $t$. The object that I use for $M_{it,t^*}$ is $prob(M_{it} = 1|M_{it}^{PSID})$, described in equation (35). The formula that I use to define health status at time $t$ is

\[
M_{it} = \frac{1}{2} M_{it,t^*} + \frac{1}{2} \left( \frac{t^*}{12} M_{it-1,t^*} + \frac{12 - t^*}{12} M_{it+1,t^*} \right) \tag{37}
\]

When determining this year’s expected level of health, this year’s health measure receives half the weight, with the other half of the weight going to both last year’s and next year’s health measure. If this year’s interview was early in the year, next year’s level of health will receive a relatively small weight and last year’s level will receive a relatively large weight.
Addressing the misreporting problem and the timing problem results have fairly large
effects on the estimated effect of health on labor force participation rates and hours worked.
Accounting for these problems in equation (16) approximately doubles the estimated effect
of health.

An alternative technique yields similar results. In this alternative technique I drop obser-
vations where the individual changed health states in nearby time periods when estimating
the fixed effects profiles in equation (16). If an individual is observed being healthy several
time periods then unhealthy for several time periods, I drop the two years before and after
the health change. This means the effect of health is identified only by respondents who were
healthy for several years then unhealthy for several years.

Appendix E: Moment Conditions

In this appendix I describe the GMM minimization procedure where I account for the three
data problems discussed in the text. The first data problem is that I wish to match profiles
that are uncontaminated by cohort and family size effects. The second problem is a selection
problem. If individuals who are healthier have a greater preference for work than unhealthy
individuals, then selection into the healthy hours and participation moment conditions will
not be random. Failure to overcome this problem will lead to an overestimate of the effect
of health on preferences for work. Healthy workers will work more hours than unhealthy
workers, not because of health but unobserved differences in preferences for leisure between
healthy and unhealthy individuals. The third data problem is that I use an unbalanced panel
of data. Since not all individuals are seen in all moment conditions during all time periods,
some of the individual level contributions to the moment conditions are “missing”. Moreover,
because individuals are healthy only with a certain probability, many of the individual level
contributions are “missing” with a certain probability.

I now discuss the first two problems and their solution in greater detail. For concrete-
ess, consider the moment condition for hours worked for individuals in good health. Upon
estimating the fixed-effects profile for hours worked, I use the estimated parameters for age
and the person-specific residual from estimation of (16). I use these estimates to generate a
predicted life cycle profile for hours worked. This life cycle profile helps me generate my set of moment conditions. I wish to set the following moment condition to zero:

$$ E \left[ \ln H_{it,M=good|\text{birthyear} = 1940, M = good, \text{famsize} = 3} - \ln \tilde{H}_{it,M=good} \right] = 0 \quad (38) $$

where $\ln \tilde{H}_{M,t}$ is the simulated geometric mean of log hours worked. In order to generate this moment condition, I use parameter estimates from equation (16) and make three modifications to hours worked, shown in equation (39):

$$ \ln H_{it,M=good} = f_i + E[f_i|\text{birthyear} = 1940, \text{prob}(M = good) = \text{prob}(M = good|\text{age} = 50), \text{age}_{it} = 50] - E[f_i|\text{birthyear}_i, \text{prob}(M_{it} = good), \text{age}_{it}] + \Pi \text{age}_{it} + \Pi f(\text{famsize} = 3) + u_{it} \quad (39) $$

Note the three modifications to the data. First, there is no probability of being in good or bad health. Instead, individuals are in either good or bad health for certain. Second, it is not the size of the family that is used but a family size of three. In this way life cycle profiles will not be contaminated by family size effects. Third, I adjust the person specific effects $f_i$. There are two things that I wish to adjust in the person specific effects. First, I wish to control for cohort effects. Note that an individual’s cohort effect is one component of their person specific effect, as a cohort effect is an average of the fixed effects of everyone born in that cohort. I adjust the person specific effect so that everyone has the same cohort effect, set to $\text{birthyear} = 1940$, which means profiles will be uncontaminated by cohort effects. The second aspect of the person specific effect that I adjust for is the possible correlation between the person specific effect and the health status of the individual. This solves the problem of selection into a moment condition, as the adjusted hours data in (39) should be uncorrelated with health.

In order to generate the adjusted hours data in (39) it is necessary to predict the person specific effect $f_i$ for a given age, birthyear, and health status. To predict the person specific fixed effect I estimate the conditional expectation of $f_i$ given birthyear and age interacted
with health status using OLS:

\[ f_i = \pi_1 \text{birthyear} + \pi_2 \text{prob}(M_{it} = \text{good}) \times \text{age}_{it} + \pi_3 (1 - \text{prob}(M_{it} = \text{good})) \times \text{age}_{it} + \epsilon_{it} \] (40)

where \( \pi_1, \pi_2, \pi_3 \) are parameters to be estimated and \( \text{birthyear} \) is a full set of birthyear dummies, and \( \text{age}_{it} \) denotes a full set of age dummies.

Next I address the problem of having an unbalanced panel and the problem of not knowing an individual’s health status with certainty. If there are \( I \) separate individuals in the data there will be a total of \( I \) possible contributions to both the healthy and unhealthy moment conditions for hours at age \( t \). However, not all individuals are observed working for all possible time periods. Assume instead that there are \( I_t \leq I \) individuals observed working at age \( t \). The idea is to treat a moment contribution if it is missing as if the contribution to the moment condition were zero.

This means that the moment condition for individuals of age \( t \) and health state \( M = \text{good} \) is generated by

\[
\frac{1}{I} \sum_{i=1}^{I_t} \left\{ \ln H_{it,M=\text{good}} - \ln \hat{H}_{it,M=\text{good}} \right\} \times \text{prob}(M_{it} = \text{good})
\] (41)

where \( \ln H_{it,M=\text{good}} \) is adjusted work hours described in equation (39). The relative weight of this moment condition rises as \( I_t \), the number of observed workers rises and as the probability that these workers are healthy rises. It is here that we can see that \( \text{prob}(M_{it} = \text{good}) \), which determines selection into the moment condition, might be correlated with the person specific fixed effect \( f_i \) but will not be correlated with its adjusted value \( f_i + E[f_i | \text{birthyear} = 1930, \text{prob}(M = \text{good}) = \text{prob}(M = \text{good} | \text{age} = 50), \text{age}_{it} = 50] - E[f_i | \text{birthyear}_i, \text{prob}(M_{it} = \text{good}), \text{age}_{it}] \) by construction.

The value of \( \theta \) that minimizes the (weighted) distance between the true profiles and estimated profiles for assets, hours, and participation is considered to be the true value of \( \theta \). Define the vector of the \( 5T \) moment conditions as \( g(\theta; \chi) \). Assuming \( W_T \) is an optimal
weighting matrix, the minimized GMM criterion function

\[
\frac{I}{1 + \tau} \bar{g}(\theta; \chi)' W_T \bar{g}(\theta; \chi)
\]  

(42)

is distributed asymptotically as Chi-squared with \(5T - 7\) degrees of freedom if the model is correctly specified. \(\tau\) 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. My estimate of \(W_T\) is the inverse of the \(5T \times 5T\) variance covariance matrix of the (adjusted) data. That is, \(W_T^{-1}\) has a typical element along the diagonal of a variance \(\frac{1}{T} \sum_{i=1}^{I_t} \{[\ln H_{it,M=good} - E[\ln H_{it,M=good}] \times \text{prob}(M_{it} = \text{good})] \times \text{prob}(M_{it} = \text{good})\}^2\) and a typical element of a covariance on the off-diagonal. When computing the chi-square statistic and the standard errors, the estimated value of \(E[\ln H_{it,M=good}]\) is replaced with its simulated counterpart.

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. Denoting \(\theta_0\) as the true parameter vector the estimated value of \(\theta_0, \hat{\theta}\), converges in distribution to

\[
\sqrt{T}(\theta - \theta_0) \overset{d}{\sim} N(0, V),
\]  

(43)

where \(V\) is the variance-covariance matrix of \(\hat{\theta}\) which is estimated by:

\[
\hat{V} = (1 + \tau)(D'W\hat{D})^{-1}
\]  

(44)
\[ \dot{D} = \frac{\partial \tilde{g}}{\partial \theta} \bigg|_{\theta = \tilde{\theta}} . \]