

SUPPLEMENT TO “THE EFFECTS OF HEALTH INSURANCE AND
SELF-INSURANCE ON RETIREMENT BEHAVIOR”
(*Econometrica*, Vol. 79, No. 3, May 2011, 693–732)

BY ERIC FRENCH AND JOHN BAILEY JONES

APPENDIX B: 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. We use the standard deduction, and thus do not allow individuals to defer medical expenses as an itemized 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 B.I).

TABLE B.I
AFTER-TAX INCOME

Pre-Tax Income (Y)	Post-Tax Income	Marginal Tax Rate
0–6,250	$0.9235Y$	0.0765
6,250–40,200	$5,771.88 + 0.7384(Y - 6,250)$	0.2616
40,200–68,400	$30,840.56 + 0.5881(Y - 40,200)$	0.4119
68,400–93,950	$47,424.98 + 0.6501(Y - 68,400)$	0.3499
93,950–148,250	$64,035.03 + 0.6166(Y - 93,950)$	0.3834
148,250–284,700	$97,515.41 + 0.5640(Y - 148,250)$	0.4360
284,700+	$174,474.21 + 0.5239(Y - 284,700)$	0.4761

APPENDIX C: COMPUTATION OF AIME

We model several key aspects of Social Security benefits. First, Social Security benefits are based on the individual’s 35 highest earnings years, relative to average wages in the economy during those years. The average earnings over these 35 highest earnings years are called Average Indexed Monthly Earnings (AIME). It immediately follows that working an additional year increases the AIME of an individual with less than 35 years of work. If an individual has already worked 35 years, he can still increase his AIME by working an additional year, but only if his current earnings are higher than the lowest earnings embedded in his current AIME. To account for real wage growth, earnings in earlier years are inflated by the growth rate of average earnings in the overall

economy. For the period 1992–1999, average real wage growth, g , was 0.016 (Committee on Ways and Means (2000, p. 923)). This indexing stops at the year the worker turns 60, however, and earnings accrued after age 60 are not rescaled.¹ Furthermore, AIME is capped. In 1998, the base year for the analysis, the maximum AIME level was \$68,400.

Precisely modelling these mechanics would require us to keep track of a worker's entire earnings history, which is computationally infeasible. As an approximation, we assume that (for workers beneath the maximum) annualized AIME is given by

$$\begin{aligned} \text{AIME}_{t+1} &= (1 + g \times 1\{t \leq 60\})\text{AIME}_t \\ &+ \frac{1}{35} \max\{0, W_t N_t - \alpha_t (1 + g \times 1\{t \leq 60\})\text{AIME}_t\}, \end{aligned}$$

where the parameter α_t approximates the ratio of the lowest earnings year to AIME. We assume that 20% of the workers enter the labor force each year between ages 21 and 25, so that $\alpha_t = 0$ for workers aged 55 and younger. For workers aged 60 and older, earnings update AIME_t only if current earnings replace the lowest year of earnings, so we estimate α_t by simulating wage (not earnings) histories with the model developed in French (2005), calculating the sequence $\{1\{\text{time-}t \text{ earnings do not increase } \text{AIME}_t\}\}_{t \geq 60}$ for each simulated wage history and estimating α_t as the average of this indicator at each age. Linear interpolation yields $\alpha_{56} - \alpha_{59}$.

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

$$(C.1) \quad \text{PIA}_t = \begin{cases} 0.9 \times \text{AIME}_t, & \text{if } \text{AIME}_t < \$5,724, \\ \$5,151.6 + 0.32 \times (\text{AIME}_t - 5,724), & \text{if } \$5,724 \leq \text{AIME}_t < \$34,500, \\ \$14,359.9 + 0.15 \times (\text{AIME}_t - 34,500), & \text{if } \text{AIME}_t \geq \$34,500. \end{cases}$$

Social Security benefits ss_t depend both on 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 that the individual first draws benefits, benefits are reduced by 6.67% and for every year (up until age 70) that benefit receipt is delayed, benefits increase by 5.0%. The effects of early or late application can be modelled as changes in AIME rather than changes in PIA, eliminating the need to include age at application as a state variable. For example, if an individual begins drawing benefits at age 62, his adjusted AIME must result in a PIA that is only 80%

¹After age 62, nominal benefits increase at the rate of inflation.

of the PIA he would have received had he first drawn benefits at age 65. Using equation (C.1), this is easy to find.

APPENDIX D: PENSIONS

Although the HRS pension data allow us to estimate pension wealth with a high degree of precision, Bellman's curse of dimensionality prevents us from including in our dynamic programming model the full range of pension heterogeneity found in the data. Thus we use the pension data to construct a simpler model of pensions. The fundamental equation behind our model of pensions is the accumulation equation for pension wealth, pw_t ,

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

where $pacc_t$ is pension accrual and pb_t is pension benefits. Two features of this equation bear noting. First, a pension is worthless once an individual dies. Dividing through by the survival probability s_{t+1} ensures that the expected value of pensions $E(pw_{t+1}|pw_t, pacc_t, pb_t)$ equals $(1+r)pw_t + pacc_t - pb_t$, the actuarially fair amount. 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.

Simulating equation (D.1) requires us to know pension benefits and pension accrual. We calculate pension benefits by assuming that at age t , the pension benefit is

$$(D.2) \quad pb_t = pf_t \times pb_t^{\max},$$

where pb_t^{\max} is the benefit received by individuals who are actually receiving pensions (given the earnings history observed at time t) and pf_t is the probability that a person with a pension is currently drawing pension benefits. We estimate pf_t as the fraction of respondents who are covered by a pension that receive pension benefits at each age; the fraction increases fairly smoothly, except for a 23-percentage-point jump at age 62. To find the annuity pb_t^{\max} given pension wealth at time t (and assuming no further pension accruals so that $pacc_k = 0$ for $k = t, t+1, \dots, T$), note first that recursively substituting equation (D.1) and imposing $pw_{T+1} = 0$ reveals that pension wealth is equal to the present discounted value of future pension benefits,

$$pw_t = \frac{1}{1+r} \sum_{k=t}^T \frac{S(k, t)}{(1+r)^{k-t}} pf_k pb_k^{\max},$$

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 the maxi-

mum pension benefit is constant from time t forward, so that $\text{pb}_k^{\max} = \text{pb}_t^{\max}$, $k = t, t + 1, \dots, T$, this equation reduces to

$$(D.3) \quad \text{pw}_t = \Gamma_t \text{pb}_t^{\max},$$

$$\Gamma_t \equiv \frac{1}{1+r} \sum_{k=t}^T \frac{S(k, t)}{(1+r)^{k-t}} \text{pf}_k.$$

Using equations (D.2) and (D.3), pension benefits are thus given by

$$(D.4) \quad \text{pb}_t = \text{pf}_t \Gamma_t^{-1} \text{pw}_t.$$

Next, we assume pension accrual is given by

$$(D.5) \quad \text{pacc}_t = \alpha_0(I_t, W_t N_t, t) \times W_t N_t,$$

where $\alpha_0(\cdot)$ is the pension accrual rate as a function of health insurance type, labor income, and age. We estimate $\alpha_0(\cdot)$ in two steps, estimating separately each component of

$$\alpha_0(I_t, W_t N_t, t) = E(\text{pacc}_t | W_t N_t, I_t, t, \text{pen}_t = 1) \\ \times \Pr(\text{pen}_t = 1 | I_t, W_t N_t),$$

where pacc_t is the accrual rate for those who have pension and pen_t is a 0–1 indicator equal to 1 if the individual has a pension.

We estimate the first component, $E(\text{pacc}_t | W_t N_t, I_t, t, \text{pen}_t = 1)$, from restricted HRS pension data. To generate a pension accrual rate for each individual, we combine the pension data with the HRS pension calculator to estimate the pension wealth that each individual would have if he left his job at different ages. The increase in pension wealth gained by working 1 more year is the accrual. Assuming that pension benefits are 0 as long as the worker continues working, it follows from equation (D.1) that

$$\text{pacc}_t = s_{t+1} \text{pw}_{t+1} - (1+r) \text{pw}_t.$$

The HRS pension data have a high degree of employer- and worker-level detail, allowing us to estimate pension accrual accurately. With accruals in hand, we then estimate $E(\text{pacc}_t | W_t N_t, I_t, t, \text{pen}_t = 1)$ by regressing accrual rates on a fourth-order age polynomial, indicators for age greater than 62 or 65, log income, log income interacted with the age variables, health insurance indicators, and health insurance indicators interacted with the age variables, using the subset of workers who have a pension on their current job.

Figure D.1 shows estimated pension accrual by health insurance type and earnings. It shows that those who have retiree coverage have the sharpest

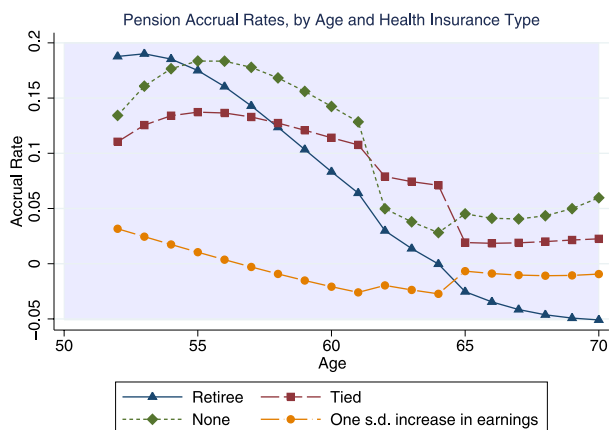


FIGURE D.1.—Pension accrual rates for individuals with pensions, by age, health insurance coverage, and earnings.

declines in pension accrual after age 60. It also shows that once health insurance and the probability of having a pension plan are accounted for, the effect of income on pension accrual is relatively small. Our estimated age (but not health insurance) pension accrual rates line up closely with Gustman, Mitchell, Samwick, and Steinmeier (2000), who also used the restricted firm-based HRS pension data.

In the second step, we estimate the probability of having a pension, $\Pr(\text{pen}_i = 1 | I_i, W_i N_i, t)$, using unrestricted self-reported data from individuals who are working and are ages 51–55. The function $\Pr(\text{pen}_i = 1 | I_i, W_i N_i, t)$ is estimated as a logistic function of log income, health insurance indicators, and interactions between log income and health insurance.

Table D.I shows the probability of having different types of pensions, conditional on health insurance. The table shows that only 8% of men with no health

TABLE D.I

PROBABILITY OF HAVING A PENSION ON THE CURRENT JOB, BY HEALTH INSURANCE TYPE, WORKING MEN, AGE 51–55

Variable	Probability of Pension Type		
	No Insurance	Retiree Insurance	Tied Insurance
Defined benefit	0.026	0.412	0.260
Defined contribution	0.050	0.172	0.270
Both DB and DC	0.006	0.160	0.106
Total	0.082	0.744	0.636
Number of observations	343	955	369

insurance have a pension, but 64% of men with tied coverage and 74% of men with retiree insurance have a pension. Furthermore, it shows that those with retiree coverage are also the most likely to have defined benefit (DB) pension plans, which provide the strongest retirement incentives after age 62.

Combining the restricted data with the HRS pension calculator also yields initial pension balances as of 1992. Mean pension wealth in our estimation sample is \$93,300. Disaggregating by health insurance type, those with retiree coverage have \$129,200, those with tied coverage have \$80,000, and those with none have \$17,300. With these starting values, we simulate pension wealth in our dynamic programming model with equation (D.1), using equation (D.5) to estimate pension accrual and using equation (D.4) to estimate pension benefits. Using these equations, it is straightforward to track and record the pension balances of each simulated individual.

Even though it is straightforward to use equation (D.1) 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 Social Security benefits:

$$\begin{aligned}
 \text{(D.6)} \quad \widehat{\text{pb}}_t(\text{PIA}_t, I_{t-1}, t) &= \sum_{k \in \{\text{retiree, tied, none}\}} [\gamma_{0,k,0} + \gamma_{0,k,1}t + \gamma_{0,k,2}t^2] \cdot 1\{I_{t-1} = k\} \\
 &\quad + \gamma_3 \text{PIA}_t + [\gamma_{4,0} + \gamma_{4,1}t + \gamma_{4,2}t^2] \cdot \max\{0, \text{PIA}_t - 9,999.6\} \\
 &\quad + [\gamma_{5,0} + \gamma_{5,1}t + \gamma_{5,2}t^2] \cdot \max\{0, \text{PIA}_t - 14,359.9\},
 \end{aligned}$$

where PIA_t is the Social Security benefit the worker would get if he were drawing benefits at time t ; as shown in Appendix C above, PIA is a monotonic function of AIME. Using equations (D.3) and (D.6) yields imputed pension wealth, $\widehat{\text{pw}}_t = \Gamma_t \widehat{\text{pb}}_t$. Equation (D.6) is estimated with regressions on simulated data generated by the model. Since these simulated data depend on the γ 's— $\widehat{\text{pw}}_t$ affects the decision rules used in the simulations—the γ 's solve a fixed-point problem. Fortunately, estimates of the γ 's converge after a few iterations.

This imputation process raises two complications. The first is that we use a different pension wealth imputation formula when calculating decision rules than we do in the simulations. If an individual's time- t pension wealth is $\widehat{\text{pw}}_t$, his time- $t + 1$ pension wealth (if living) should be

$$\widehat{\text{pw}}_{t+1} = (1/s_{t+1})[(1+r)\widehat{\text{pw}}_t + \text{pacc}_t - \text{pb}_t].$$

This quantity, however, might differ from the pension wealth that would be imputed using PIA_{t+1} , $\widehat{\text{pw}}_{t+1} = \Gamma_{t+1} \widehat{\text{pb}}_{t+1}$, where $\widehat{\text{pb}}_{t+1}$ is defined in equation (D.6).

To correct for this, we increase nonpension wealth, A_{t+1} , by $s_{t+1}(1 - \tau_t)(\widehat{pw}_{t+1} - \widehat{pw}_{t+1})$. The first term in this expression reflects the fact that while nonpension 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 nonpension 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 annual Social Security benefits, it should not affect his pension benefits. (Recall that we reduce PIA if an individual draws benefits before age 65.) The pension imputation procedure we use, however, would imply that it does. We counter this problem by recalculating PIA when the individual begins drawing Social Security benefits. In particular, suppose that a decision to accelerate or defer application changes PIA_t to $rem_t PIA_t$. Our approach is to use equation (D.6) find a value PIA_t^* such that

$$(1 - \tau_t)\widehat{pb}_t(PIA_t^*) + PIA_t^* = (1 - \tau_t)\widehat{pb}_t(PIA_t) + rem_t PIA_t,$$

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

APPENDIX E: 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 backward to time 1. We find the time- T decisions by solving the time- T Bellman equation at each value of the state vector X_T , with the terminal value function set equal to bequest utility: $V_{T+1} = b(A_{T+1})$. 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 the time- $(T - 1)$ Bellman equation, having solved for V_T already. Continuing this backward induction yields decision rules for times $T - 2, T - 3, \dots, 1$.

The value function is directly computed at a finite number of points within a grid, $\{X_i\}_{i=1}^I$.² We use linear interpolation within the grid (i.e., we take a weighted average of the value functions of the surrounding grid points) 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

²In practice, the grid consists of 32 asset states, $A_h \in [-\$55,000, \$1,200,000]$; 5 wage residual states, $\omega_i \in [-0.99, 0.99]$; 16 AIME states, $AIME_j \in [\$4,000, \$68,400]$; 3 states for the persistent component of medical expenses, ζ_k , over a normalized (unit variance) interval of $[-1.5, 1.5]$. There are also two application states, two health states, and two states for participation in the previous period. This requires solving the value function at 61,440 different points for ages 62–69, when the individual is eligible to apply for benefits, at 31,260 points before age 62 (when application is not an option), or at ages 70–71 (when we impose application), and at 15,360 points after age 71 (when we impose retirement as well).

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 ζ_t and ω_{t+1} into discrete Markov chains, following the approach of Tauchen (1986); using discretization rather than quadrature greatly reduces the number of times one has to interpolate when calculating $E_t(V(X_{t+1}))$. We integrate the value function with respect to the transitory component of medical expenses, ξ_t , using five-node Gauss–Hermite quadrature (see Judd (1998)).

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.³

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 histories. Given the realized state vector X_{i0} , individual i ’s realized decisions at time 0 are found by evaluating the time-0 decision functions at X_{i0} . Using the asset accumulation equation and budget constraints described in the main text, we combine X_{i0} , the time-0 decisions, and the individual i ’s time-1 shocks to get the time-1 state vector, X_{i1} . 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 the shocks ζ_t and ω_t as well. While these processes are approximated as finite Markov chains when the decision rules are found, the simulated sequences of ζ_t and ω_t are generated from continuous processes. This makes the simulated life-cycle profiles less sensitive to the discretization of ζ_t and ω_t than when ζ_t and ω_t are drawn from Markov chains.

³The consumption grid has 100 points, and the hours grid is broken into 500-hour intervals. When this grid is used, the consumption search at a value of the state vector X for time t is centered around the consumption grid point that was optimal for the same value of X at time $t + 1$. (Recall that we solve the model backward in time.) 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 six 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 $N_t = 0$, searching around the consumption choice that was optimal when $H_{t+1} = 0$. Once these values are found, we perform a quick, “second-pass” search in a neighborhood around them.

Finally, to reduce the computational burden, we assume that all workers apply for Social Security benefits by age 70 and retire by age 72: for $t \geq 70$, $B_t = 1$, and for $t \geq 72$, $N_t = 0$.

APPENDIX F: MOMENT CONDITIONS, ESTIMATION MECHANICS, AND THE ASYMPTOTIC DISTRIBUTION OF PARAMETER ESTIMATES

Following Gourinchas and Parker (2002), French (2005), and Laibson, Repetto, and Tobacman (2007), we estimate the parameters of the model in two steps. In the first step we estimate or calibrate parameters that can be cleanly identified without explicitly using our model. For example, we estimate mortality rates and health transitions from demographic data. As a matter of notation, we call this set of parameters χ . In the second step, we estimate the vector of “preference” parameters, $\theta = (\gamma_0, \gamma_1, \gamma_2, \beta_0, \beta_1, \beta_2, \nu, L, \phi_{P0}, \phi_{P1}, \phi_{RE}, \theta_B, \kappa, C_{\min}, \text{preference type prediction coefficients})$, using the method of simulated moments (MSM).

We assume that the “true” preference vector θ_0 lies in the interior of the compact set $\Theta \subset \mathbb{R}^{39}$. Our estimate, $\hat{\theta}$, is the value of θ 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 $21T$ moment conditions. They are, for each age $t \in \{1, \dots, T\}$, two asset quantiles (forming $2T$ moment conditions), labor force participation rates conditional on asset quantile and health insurance type ($9T$), labor market exit rates for each health insurance type ($3T$), labor force participation rates conditional on the preference indicator described in the main text ($3T$), and labor force participation rates and mean hours worked conditional on health status ($4T$).

Consider first the asset quantiles. As stated in the main text, let $j \in \{1, 2, \dots, J\}$ index asset quantiles, where J is the total number of asset quantiles. Assuming that the age-conditional distribution of assets is continuous, the π_j th age-conditional asset quantile, $Q_{\pi_j}(A_{it}, t)$, is defined as

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

In other words, the fraction of age- t individuals with less than Q_{π_j} in assets is π_j . As is well known (see, e.g., Manski (1988), Powell (1994), and Buchinsky (1998) or the review in Chernozhukov and Hansen (2002)), the preceding equation can be rewritten as a moment condition by using the indicator function

$$(F.1) \quad E(1\{A_{it} \leq Q_{\pi_j}(A_{it}, t)\} | t) = \pi_j.$$

The model analog to $Q_{\pi_j}(A_{it}, t)$ is $g_{\pi_j}(t; \theta_0, \chi_0)$, the j th quantile of the simulated asset distribution. If the model is true, then the data quantile in equation (F.1) can be replaced by the model quantile, and equation (F.1) can be

rewritten as

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

Since $J = 2$, equation (F.2) generates $2T$ moment conditions.

Equation (F.2) is a departure from the usual practice of minimizing a sum of weighted absolute errors in quantile estimation. The quantile restrictions just described, however, are part of a larger set of moment conditions, which means that we can no longer estimate θ by minimizing weighted absolute errors. Our approach to handling multiple quantiles is similar to the minimum distance framework used by Epple and Seig (1999).⁴

The next set of moment conditions uses the quantile-conditional means of labor force participation. Let $\bar{P}_j(I, t; \theta_0, \chi_0)$ denote the model's prediction of labor force participation given asset quantile interval j , health insurance type I , and age t . If the model is true, $\bar{P}_j(I, t; \theta_0, \chi_0)$ should equal the conditional participation rates found in the data,

$$\bar{P}_j(I, t; \theta_0, \chi_0) = E[P_{it} | I, t, g_{\pi_{j-1}}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0)]$$

with $\pi_0 = 0$ and $\pi_{J+1} = 1$. Using indicator function notation, we can convert this conditional moment equation into an unconditional one (e.g., Chamberlain (1992)):

$$(F.3) \quad E([P_{it} - \bar{P}_j(I, t; \theta_0, \chi_0)] \times 1\{I_{it} = I\} \\ \times 1\{g_{\pi_{j-1}}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0)\} | t) = 0$$

for $j \in \{1, 2, \dots, J + 1\}$, $I \in \{\text{none, retiree, tied}\}$, $t \in \{1, \dots, T\}$. Note that $g_{\pi_0}(t) \equiv -\infty$ and $g_{\pi_{J+1}}(t) \equiv \infty$. With two quantiles (generating three quantile-conditional means) and three health insurance types, equation (F.3) generates $9T$ moment conditions.

As described in Appendix J, we use HRS attitudinal questions to construct the preference index $\text{pref} \in \{\text{high, low, out}\}$. Considering how participation varies across this index leads to the moment condition

$$(F.4) \quad E(P_{it} - \bar{P}(\text{pref}, t; \theta_0, \chi_0) | \text{pref}_i = \text{pref}, t) = 0$$

for $t \in \{1, \dots, T\}$, $\text{pref} \in \{0, 1, 2\}$. Equation (F.4) yields $3T$ moment conditions, which are converted into unconditional moment equations with indicator functions.

⁴Buchinsky (1998) showed that one could include the first-order conditions from multiple absolute value minimization problems in the moment set. However, his approach involves finding the gradient of $g_{\pi_j}(t; \theta, \chi)$ at each step of the minimization search.

We also match exit rates for each health insurance category. Let $\overline{EX}(I, t; \theta_0, \chi_0)$ denote the fraction of time- $t - 1$ workers predicted to leave the labor market at time t . The associated moment condition is

$$(F.5) \quad E([1 - P_{it}] - \overline{EX}(I, t; \theta_0, \chi_0) | I_{i,60} = I, P_{i,t-1} = 1, t) = 0$$

for $I \in \{\text{none, retiree, tied}\}$, $t \in \{1, \dots, T\}$. Equation (F.5) generates $3T$ moment conditions, which are converted into unconditional moments as well.⁵

Finally, consider health-conditional hours and participation. Let $\overline{\ln N}(H, t; \theta_0, \chi_0)$ and $\overline{P}(H, t; \theta_0, \chi_0)$ denote the conditional expectation functions for hours (when working) and participation generated by the model for workers with health status H ; let $\ln N_{it}$ and P_{it} denote measured hours and participation. The moment conditions are

$$(F.6) \quad E(\ln N_{it} - \overline{\ln N}(H, t; \theta_0, \chi_0) | P_{it} > 0, H_{it} = H, t) = 0,$$

$$(F.7) \quad E(P_{it} - \overline{P}(H, t; \theta_0, \chi_0) | H_{it} = H, t) = 0$$

for $t \in \{1, \dots, T\}$, $H \in \{0, 1\}$. Equations (F.6) and (F.7), once again converted into unconditional form, yield $4T$ moment conditions, for a grand total of $21T$ moment conditions.

Combining all the moment conditions described here is straightforward: we simply stack the moment conditions and estimate jointly.

Suppose we have a data set of I independent individuals who are each observed for T periods. Let $\varphi(\theta; \chi_0)$ denote the $21T$ -element vector of moment conditions that was described in the main text and immediately above, and let $\hat{\varphi}_I(\cdot)$ denote its sample analog. 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 $\varphi(\cdot)$ to be "missing," as in French and Jones (2004). Letting $\widehat{\mathbf{W}}_I$ denote a $21T \times 21T$ weighting matrix, the MSM estimator $\hat{\theta}$ is given by

$$(F.8) \quad \arg \min_{\theta} \frac{I}{1 + \tau} \hat{\varphi}_I(\theta, \chi_0)' \widehat{\mathbf{W}}_I \hat{\varphi}_I(\theta, \chi_0),$$

where τ is the ratio of the number of observations to the number of simulated observations.

To find the solution to equation (F.8), we proceed as follows:

⁵Because exit rates apply only to those working in the previous period, they normally do not contain the same information as participation rates. However, this is not the case for workers with tied coverage, as a worker stays in the tied category only as long as he continues to work. To remove this redundancy, the exit rates in equation (F.5) are conditioned on the individual's age-60 health insurance coverage, while the participation rates in equation (F.3) are conditioned on the individual's current coverage.

Step 1. We aggregate the sample data into life-cycle profiles for hours, participation, exit rates, and assets.

Step 2. 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. See Appendix G for details. We also use these data to estimate many of the parameters contained in the belief vector χ , although we calibrate other parameters. The initial distribution also includes preference type, assigned using our type prediction equation.

Step 3. Using χ , we generate matrices of random health, wage, mortality, and medical expense shocks. The matrices hold shocks for 90,000 simulated individuals.

Step 4. We compute the decision rules for an initial guess of the parameter vector θ , using χ and the numerical methods described in Appendix E.

Step 5. We simulate profiles for the decision variables. Each simulated individual receives a draw of preference type, assets, health, wages, and medical expenses from the initial distribution, and is assigned one of the simulated sequences of health, wage, and medical expense 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 the next period's states, and the simulated shocks pin the states down exactly.

Step 6. We aggregate the simulated data into life-cycle profiles.

Step 7. We compute moment conditions, that is, we find the distance between the simulated and true profiles, as described in equation (F.8).

Step 8. We pick a new value of θ , update the simulated distribution of preference types, and repeat Steps 4–7 until we find the value of θ that minimizes the distance between the true data and the simulated data. This vector of parameter values, $\hat{\theta}$, is our estimated value of θ_0 .⁶

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:

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

with the variance–covariance matrix \mathbf{V} given by

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

where \mathbf{S} is the $21T \times 21T$ variance–covariance matrix of the data,

$$(F.9) \quad \mathbf{D} = \left. \frac{\partial \varphi(\theta, \chi_0)}{\partial \theta'} \right|_{\theta=\theta_0}$$

⁶Because the GMM criterion function is discontinuous, we search over the parameter space using a Simplex algorithm written by Honore and Kyriazidou. It usually takes 2–4 weeks to estimate the model on a 48-node cluster, with each iteration (of Steps 4–7) taking around 15 minutes.

is the $21T \times 39$ Jacobian matrix of the population moment vector, and $\mathbf{W} = \text{plim}_{T \rightarrow \infty} \{\widehat{\mathbf{W}}_T\}$. Moreover, Newey (1985) showed that if the model is properly specified,

$$\frac{I}{1 + \tau} \widehat{\boldsymbol{\varphi}}_T(\hat{\boldsymbol{\theta}}, \chi_0)' \mathbf{R}^{-1} \widehat{\boldsymbol{\varphi}}_T(\hat{\boldsymbol{\theta}}, \chi_0) \rightsquigarrow \chi_{21T-39}^2,$$

where \mathbf{R}^{-1} is the generalized inverse of

$$\mathbf{R} = \mathbf{PSP},$$

$$\mathbf{P} = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}.$$

The asymptotically efficient weighting matrix arises when $\widehat{\mathbf{W}}_T$ converges to \mathbf{S}^{-1} , the inverse of the variance–covariance matrix of the data. When $\mathbf{W} = \mathbf{S}^{-1}$, \mathbf{V} simplifies to $(1 + \tau)(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$ and \mathbf{R} is replaced with \mathbf{S} . But even though the optimal weighting matrix is asymptotically efficient, it can be severely biased in small samples. (See, for example, Altonji and Segal (1996).) We thus use a “diagonal” weighting matrix, as suggested by Pischke (1995). The diagonal weighting scheme uses the inverse of the matrix that is the same as \mathbf{S} along the diagonal and has zeros off the diagonal of the matrix.

We estimate \mathbf{D} , \mathbf{S} , and \mathbf{W} with their sample analogs. For example, our estimate of \mathbf{S} is the $21T \times 21T$ estimated variance–covariance matrix of the sample data.⁷ That is, one diagonal element of $\widehat{\mathbf{S}}_T$ will be the variance estimate $\frac{1}{T} \sum_{i=1}^I [1\{A_{it} \leq Q_{\pi_j}(A_{it}, t)\} - \pi_j]^2$, while a typical off-diagonal element is a covariance. When estimating parameters, we use sample statistics, so that $Q_{\pi_j}(A_{it}, t)$ is replaced with the sample quantile $\widehat{Q}_{\pi_j}(A_{it}, t)$. When computing the chi-square statistic and the standard errors, we use model predictions, so that Q_{π_j} is replaced with its simulated counterpart, $g_{\pi_j}(t; \hat{\boldsymbol{\theta}}, \hat{\chi})$. Covariances between asset quantiles and hours and labor force participation are also simple to compute.

The gradient in equation (F.9) is straightforward to estimate for most moment conditions; we merely take numerical derivatives of $\widehat{\boldsymbol{\varphi}}_T(\cdot)$. However, in the case of the asset quantiles and quantile-conditional labor force participation, discontinuities make the function $\widehat{\boldsymbol{\varphi}}_T(\cdot)$ nondifferentiable at certain data points. Therefore, our results do not follow from the standard GMM approach, but rather the approach for nonsmooth functions described in Pakes and Pollard (1989), Newey and McFadden (1994, Section 7), and Powell (1994). We find the asset quantile component of \mathbf{D} by rewriting equation (F.2) as

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

⁷In one asset tertile-age-insurance type cell, all workers make the same participation decision. Rather than discard the cell, we smooth across these moment conditions to ensure that the variance–covariance matrix \mathbf{S} is nonsingular. The need for such adjustments should vanish as the sample size grows.

where $F(g_{\pi_j}(t; \theta_0, \chi_0)|t)$ is the empirical cumulative distribution function of time- t assets evaluated at the model-predicted π_j th quantile. Differentiating this equation yields

$$\mathbf{D}_{jt} = f(g_{\pi_j}(t; \theta_0, \chi_0)|t) \frac{\partial g_{\pi_j}(t; \theta_0, \chi_0)}{\partial \theta'}$$

where \mathbf{D}_{jt} is the row of \mathbf{D} corresponding to the π_j th quantile at year t . In practice, we find $f(g_{\pi_j}(t; \theta_0, \chi_0)|t)$, the probability density function of time- t assets evaluated at the π_j th quantile, with a kernel density estimator. We use a kernel estimator for GAUSS written by Ruud Koning.

To find the component of the matrix \mathbf{D} for the asset-conditional labor force participation rates, it is helpful to write equation (F.3) as

$$\begin{aligned} \Pr(I_{t-1} = I) \times \int_{g_{\pi_{j-1}}(t; \theta_0, \chi_0)}^{g_{\pi_j}(t; \theta_0, \chi_0)} [E(P_{it}|A_{it}, I, t) - \bar{P}_j(I, t; \theta_0, \chi_0)] \\ \times f(A_{it}|I, t) dA_{it} = 0, \end{aligned}$$

which implies that

$$\begin{aligned} \mathbf{D}_{jt} = & \left[-\Pr(g_{\pi_{j-1}}(t; \theta_0, \chi_0) \leq A_{it} \leq g_{\pi_j}(t; \theta_0, \chi_0)|I, t) \frac{\partial \bar{P}_j(I, t; \theta_0, \chi_0)}{\partial \theta'} \right. \\ & + [E(P_{it}|g_{\pi_j}(t; \theta_0, \chi_0), I, t) - \bar{P}_j(I, t; \theta_0, \chi_0)] \\ & \times f(g_{\pi_j}(t; \theta_0, \chi_0)|I, t) \frac{\partial g_{\pi_j}(t; \theta_0, \chi_0)}{\partial \theta'} \\ & - [E(P_{it}|g_{\pi_{j-1}}(t; \theta_0, \chi_0), I, t) - \bar{P}_j(I, t; \theta_0, \chi_0)] \\ & \left. \times f(g_{\pi_{j-1}}(t; \theta_0, \chi_0)|I, t) \frac{\partial g_{\pi_{j-1}}(t; \theta_0, \chi_0)}{\partial \theta'} \right] \\ & \times \Pr(I_{t-1} = I), \end{aligned}$$

$$\text{with } f(g_{\pi_0}(t; \theta_0, \chi_0)|I, t) \frac{\partial g_{\pi_0}(t; \theta_0, \chi_0)}{\partial \theta'} = f(g_{\pi_{J+1}}(t; \theta_0, \chi_0)|I, t) \frac{\partial g_{\pi_{J+1}}(t; \theta_0, \chi_0)}{\partial \theta'} \equiv 0.$$

APPENDIX G: DATA AND INITIAL JOINT DISTRIBUTION OF THE STATE VARIABLES

Our data are drawn from the HRS, a sample of noninstitutionalized individuals aged 51–61 in 1992. The HRS surveys individuals every 2 years; we have eight waves of data covering the period 1992–2006. We use men in the analysis.

We dropped respondents for the following reasons. First, we drop all individuals who spent over 5 years working for an employer who did not contribute to

Social Security. These individuals usually work for state governments. We drop these people because they often have very little in the way of Social Security wealth, but a great deal of pension wealth, a type of heterogeneity our model is not well suited to handle. Second, we drop respondents with missing information on health insurance, labor force participation, hours, and assets. When estimating labor force participation by asset quantile and health insurance for those born in 1931–1935 for the estimation sample [and 1936–1941 for the validation sample], we begin with 21,376 [36,702] person-year observations. We lose 3,872 [6,919] observations because of missing labor force participation, 2,109 [2,480] observations for those who worked over 5 years for firms that did not contribute to Social Security, 602 [1,074] observations due to missing wave-1 labor force participation (needed to construct the preference index), and 2,103 [3,023] observations due to missing health insurance data. In the end, from a potential sample of 21,376 [36,702] person-year observations for those between ages 51 and 69, we keep 12,870 [23,206] observations.

The labor market measures used in our analysis are constructed as follows. Hours of work are the product of usual hours per week and usual weeks per year. To compute hourly wages, we use information on how respondents are paid, how often they are paid, and how much they are paid. For salaried workers, annual earnings are 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 nonsurvey (e.g., 1993) years as missing.

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. Because we are particularly interested in labor force participation, we use the work history to construct a measure of whether the individual worked in nonsurvey years. 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.

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 nonsurvey 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. Given that wealth changes rather slowly over time, these imputations should not severely bias our results.

Medical expenses are the sum of insurance premia paid by households, drug costs, and out-of-pocket costs for hospital, nursing home care, doctor visits, dental visits, and outpatient care. As noted in the text, the proper measure of medical expenses for our model includes payments made by Medicaid. Although individuals in the HRS report whether they received Medicaid, they do not report the payments. 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 aged 65 and older, and \$9,097 per blind or disabled 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.

To measure health status we use responses to the question, “Would you say that your health is excellent, very good, good, fair, or poor?” We consider the individual in bad health if he responds “fair” or “poor,” and consider him in good health otherwise.⁸ We treat the health status for nonsurvey years as missing. Appendix H describes how we construct the health insurance indicator.

We use Social Security Administration earnings histories to construct AIME. Approximately 74% of our sample released their Social Security number to the HRS, which allowed them to be linked to their Social Security earnings histories. For those who did not release their histories, we use the procedure described below to impute AIME as a function of assets, health status, health insurance type, labor force participation, and pension type.

The HRS collects pension data from both workers and employers. The HRS asks individuals about their earnings, tenure, contributions to defined contribution (DC) plans, and their employers. HRS researchers then ask employers about the pension plans they offer their employees. If the employer offers different plans to different employees, the employee is matched to the plan based on other factors, such as union status. Given tenure, earnings, DC contributions, and pension plan descriptions, it is then possible to calculate pension wealth for each individual who reports the firm he works for. Following Scholz et al. (2006), we use firm reports of defined benefit (DB) pension wealth and individual reports of DC pension wealth if they exist. If not, we use firm-reported DC wealth and impute DB wealth as a function of wages, hours, tenure, health insurance type, whether the respondent also has a DC plan, health status, age, assets, industry, and occupation. We discuss the imputation procedure below.

Workers are asked about two different jobs: (i) their current job if working or last job if not working; (ii) the job preceding the one listed in part (i), if the individual worked at that job for over 5 years. Pension wealth from both of these jobs is included in our measure of pension wealth. Below we give descriptives for our estimation sample (born 1931–1935) and validation sample (born 1936–1941). 41% of our estimation sample [and 52% of our validation sample] are currently working and have a pension (of which 56% [57% for the validation sample] have firm-based pension details), 6% [5%] are not working and had a pension on their last job (of which 62% [62%] have firm-based pension details), and 32% [32%] of all individuals had a pension on another job (of which 35% [29%] have firm-based pension details).

To generate the initial joint distribution of assets, wages, AIME, pensions, participation, health insurance, health status, and medical expenses, we draw

⁸Bound et al. (2010) considered a more detailed measure of health status.

random vectors (i.e., random draws of individuals) from the empirical joint distribution of these variables for individuals aged 57–61 in 1992, or 1,701 observations. We drop observations with missing data on labor force participation, health status, insurance, assets, and age. We impute values for observations with missing wages, medical expenses, pension wealth, and AIME.

To impute these missing variables, we follow David, Little, Samuhel, and Triest (1986) and Little (1988), and use the following predictive mean matching regression approach. First, we regress the variable of interest y (e.g., pension wealth) on the vector of observable variables x , yielding $y = x\beta + \epsilon$. Second, for each sample member i , we calculate the predicted value $\hat{y}_i = x_i\hat{\beta}$, and for each member with an observed value of y_i , we calculate the residual $\hat{\epsilon}_i = y_i - \hat{y}_i$. Third, we sort the predicted value \hat{y}_i into deciles. Fourth, for missing observations, we impute ϵ_i by finding a random individual j with a value of \hat{y}_j in the same decile as \hat{y}_i and setting $\epsilon_i = \hat{\epsilon}_j$. The imputed value of y_i is $\hat{y}_i + \hat{\epsilon}_j$.

As David et al. (1986) pointed out, our imputation approach is equivalent to hot decking when the “ x ” variables are discretized and include a full set of interactions. The advantages of our approach over hot decking are twofold. First, many of the x variables are continuous, and it seems unwise to discretize them. Second, we have very few observations for some variables (such as pension wealth on past jobs), and hot decking is very data intensive. A small number of x variables generate a large number of hot-decking cells, as hot decking uses a full set of interactions. We found that the interaction terms are relatively unimportant, but adding extra variables was very important for improving goodness of fit when imputing pension wealth.

If someone is not working (and thus does not report a wage), we use the wage on their last job as a proxy for their current wage if it exists, and otherwise impute the log wage as a function of assets, health, health insurance type, labor force participation, AIME, and quarters of covered work. We predict medical expenses using assets, health, health insurance type, labor force participation, AIME, and quarters of covered earnings.

Last, we must infer the persistent component of the medical expense residual from the medical expenses observed in the initial distribution. Recall that the process for medical expenses is

$$(G.1) \quad \ln M_t = m(H_t, I_t, t, P_t) + \sigma(H_t, I_t, t, P_t) \times \psi_t, \\ \psi_t = \zeta_t + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2), \\ \zeta_t = \rho_m \zeta_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2).$$

Given an initial distribution of medical expenses, we calculate ζ_t , the persistent medical expense component, by first finding the normalized log deviation ψ_t , as described by equation (G.1), and then applying standard projection formulae to impute ζ_t from ψ_t .

APPENDIX H: MEASUREMENT OF HEALTH INSURANCE TYPE AND
LABOR FORCE PARTICIPATION

Much of the identification in this paper comes from differences in medical expenses and job exit rates between those with tied health insurance coverage and those with retiree coverage. Unfortunately, identifying these health insurance types is not straightforward. The HRS has rather detailed questions about health insurance, but the questions asked vary from wave to wave. Moreover, in no wave are the questions asked consistent with our definitions of tied or retiree coverage. Fortunately, our estimated health-insurance-specific job exit rates are not very sensitive to our definition of health insurance, as we show below.

In all of the HRS waves (but not AHEAD waves 1 and 2), the respondent is asked whether he has insurance provided by a current or past employer or union, or a spouse's current or past employer or union. If he responds "no" to this question, we code his coverage as none. We assume that this question is answered accurately, so that there is no measurement error when an individual reports that his insurance category is none. All of the measurement error problems arise when we allocate individuals with employer-provided coverage between the retiree and tied categories.

If an individual has employer-provided coverage in waves 1 and 2, he is asked, "Is this health insurance available to people who retire?" In waves 3–8, the analogous question is, "If you left your current employer now, could you continue this health insurance coverage up to the age of 65?" For individuals younger than 65, the question asked in waves 3–8 is a more accurate measure of whether the individual has retiree coverage. In particular, a "yes" response in waves 1 and 2 might mean only that the individual had tied coverage, but could acquire COBRA coverage if he left his job. Thus the fraction of individuals younger than 65 who report that they have employer-provided health insurance but who answer "no" to the follow-up question roughly doubles between waves 2 and 3. On the other hand, for those older than 65, the question used in waves 3–8 is meaningless.

Our preferred approach is to use the wave-1 response to determine who has retiree coverage. It is possible, however, to estimate the probability of response error to this variable. Consider first the problem of distinguishing the retiree and tied types for those younger than 65. As a matter of notation, let I denote an individual's actual health insurance coverage, and let I^* denote the measure of coverage generated by the HRS questions. To simplify the notation, assume that the individual is known to have employer-provided coverage— $I = \text{tied}$ or $I = \text{retiree}$ —so that we can drop the conditioning statement in the analysis below. Recall that many individuals who report retiree coverage in waves 1 and 2 likely have tied coverage. We are therefore interested in the misreporting probability $\Pr(I = \text{tied} | I^* = \text{retiree}, \text{wv} < 3, t < 65)$, where wv denotes HRS wave and t denotes age. To find this quantity, note first that by the law of total

probability,

$$\begin{aligned}
 \text{(H.1)} \quad & \Pr(I = \text{tied} | \text{wv} < 3, t < 65) \\
 &= \Pr(I = \text{tied} | I^* = \text{tied}, \text{wv} < 3, t < 65) \\
 &\quad \times \Pr(I^* = \text{tied} | \text{wv} < 3, t < 65) \\
 &\quad + \Pr(I = \text{tied} | I^* = \text{retiree}, \text{wv} < 3, t < 65) \\
 &\quad \times \Pr(I^* = \text{retiree} | \text{wv} < 3, t < 65).
 \end{aligned}$$

Now assume that all reports of tied coverage in waves 1 and 2 are true:

$$\Pr(I = \text{tied} | I^* = \text{tied}, \text{wv} < 3, t < 65) = 1.$$

Assume further that for individuals younger than 65 there is no measurement error in waves 3–8, and that the share of younger individuals with tied coverage is constant across waves:

$$\begin{aligned}
 \Pr(I = \text{tied} | \text{wv} < 3, t < 65) &= \Pr(I = \text{tied} | \text{wv} \geq 3, t < 65) \\
 &= \Pr(I^* = \text{tied} | \text{wv} \geq 3, t < 65).
 \end{aligned}$$

Inserting these assumptions into equation (H.1) and rearranging yields the mismeasurement probability

$$\begin{aligned}
 \text{(H.2)} \quad & \Pr(I = \text{tied} | I^* = \text{retiree}, \text{wv} < 3, t < 65) \\
 &= \frac{\Pr(I^* = \text{tied} | \text{wv} \geq 3, t < 65) - \Pr(I^* = \text{tied} | \text{wv} < 3, t < 65)}{\Pr(I^* = \text{retiree} | \text{wv} < 3, t < 65)}.
 \end{aligned}$$

To account for mismeasurement in waves 1 and 2 for those 65 and older, we again assume that all reports of tied health insurance are true. We assume further that $\Pr(I = \text{tied} | I^* = \text{retiree}, \text{wv} < 3, t \geq 65) = \Pr(I = \text{tied} | I^* = \text{retiree}, \text{wv} < 3, t < 65)$: the fraction of retiree reports in waves 1 and 2 that are inaccurate is the same across all ages. We can then apply the mismeasurement probability for people younger than 65, given by equation (H.2), to retiree reports by people 65 and older.

The second misreporting problem is that the “follow-up” question in waves 3–8 is completely uninformative for those older than 65. Our strategy for handling this problem is to treat the first observed health insurance status for these individuals as their health insurance status throughout their lives. Since we assume that reports of tied coverage are accurate, older individuals reporting tied coverage in waves 1 and 2 are assumed to receive tied coverage in waves 3–8. (Recall, however, that if an individual with tied coverage drops out of the labor market, his health insurance is none for the rest of his life.) For older individuals reporting retiree coverage in waves 1 and 2, we assume that

the misreporting probability—when we choose to account for it—is the same throughout all waves. (Recall that our preferred assumption is to assume that a “yes” response to the follow-up question in waves 1 and 2 indicates retiree coverage.)

A related problem is that individuals’ health insurance reports often change across waves, in large part because of the misreporting problems just described. Our preferred approach for handling this problem is to classify individuals on the basis of their first observed health insurance report. We also consider the approach of classifying individuals on the basis of their report from the previous wave.

Figure H.1 shows how our treatment of these measurement problems affects measured job exit rates. The top two graphs in Figure H.1 do not adjust for measurement error. The bottom two graphs account for the measurement error problems, using the approach described by equation (H.2). The two graphs in the left column use the first observed health insurance report, whereas the graphs in the right column use the previous period’s health insurance report. Figure H.1 shows that the profiles are not very sensitive to these changes. Those with retiree coverage tend to exit the labor market at age 62, whereas those with tied and no coverage tend to exit the labor market at age 65.

Another, more conceptual, problem is that 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.⁹ To circumvent this problem, we use health insurance in the previous wave and our model of health insurance transitions to predict health insurance options. For example, if in the previous wave, an individual reports working and having health insurance that is tied to his job, that individual’s choice set is either tied health insurance and working or COBRA insurance and not working.¹⁰

Our preferred specification, which we use in the analysis, is to use the first observed health insurance report and to not use the measurement error corrections.

Because agents in our model are forward-looking, we need to know the health-insurance-conditional process for medical expenses facing the very old. The data we use to estimate medical expenses for those over age 70 comes from the Assets and Health Dynamics of the Oldest Old survey. French and

⁹For example, the model predicts that all HRS respondents younger than 65 who report having tied health insurance 2 years before the survey date, work 1 year before the survey date, and are not currently working should report having COBRA coverage on the survey date. However, 19% of them report having no health insurance.

¹⁰We are assuming that everyone eligible for COBRA takes up coverage. In practice, only about $\frac{2}{3}$ of those eligible take up coverage (Gruber and Madrian (1996)). As a robustness check, we shut down the COBRA option (impose a 0% take-up rate) and reran the model. Eliminating COBRA had only a small effect on labor supply.

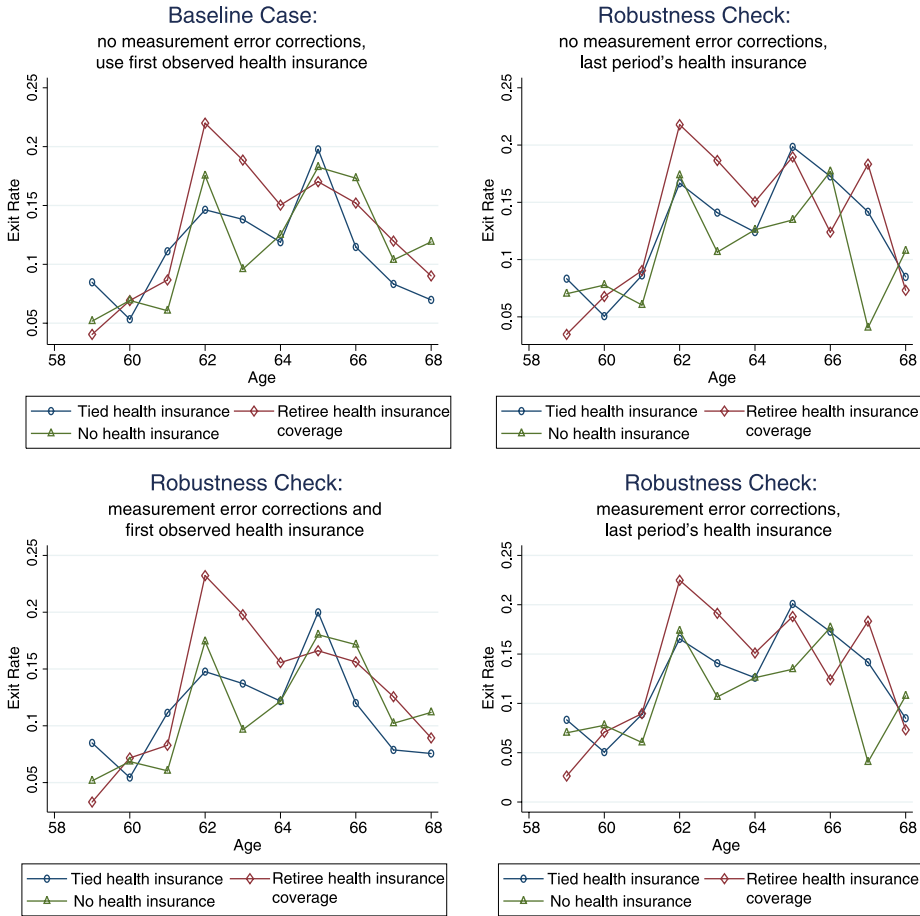


FIGURE H.1.—Job exit rates using different measures of health insurance type.

Jones (2004) discussed some of the details of the survey, as well as some of our coding decisions. The main problem with the AHEAD is that there is no question asked of respondents about whether they would lose their health insurance if they left their job, so it is not straightforward to distinguish those who have retiree coverage from those with tied coverage. To distinguish these two groups, we do the following. If the individual exits the labor market during our sample and has employer-provided health insurance at least 1 full year after exiting the labor market, we assume that individual has retiree coverage. All individuals who have employer-provided coverage when first observed, but do not meet this criterion for having retiree coverage, are assumed to have tied coverage.

Our measure of labor force participation is based on the values reported at the time of the interview. We also use the age at the time of the interview. For this reason, some of our “65-year-olds” are 65 years and 0 days old, whereas others are 65 years and 364 days old. Blau (1994) showed that most age-65 job exits occur within a few months of the 65th birthday. Thus, we may be understating the decline in labor supply at age 65, because our participation measure combines individuals who are exactly 65, who may not have yet left the labor force, with those who are almost 66, who may have left the labor force market months before.

To investigate how this timing issue affects our estimated job exit rates, we use HRS labor force histories, which provide the dates at which individuals leave the labor force, to construct three different measures of participation by age. Figure H.2 presents job exit rates derived with the different measures.

The top left panel of Figure H.2 shows job exit rates derived with the measure of participation that we use in the paper (participation at the time of the interview). In the top right panel, participation is measured at the time of the respondent’s birthday, so that the job exit rate at age 65 measures the probability that an individual was working on his 64th birthday but not on his 65th birthday. Relative to the baseline case, the peaks in exit rates at ages 62 and 65 are now less pronounced. The reason for this is that people who report leaving in the months after a 65th birthday are coded as having left at age 66. For example, an individual leaving the labor market at age 65 and 1 day would be classified as exiting the labor market at age 66. As a result, measuring labor force participation at birthdays leads to a higher estimated job exit rate at 66 and a lower rate at 65 than our baseline approach.

In the bottom left panel of Figure H.2, participation is measured at the midpoint between the respondents’ birthdays. For example, participation at age 65 is measured at age $65\frac{1}{2}$, so that the job exit rate at age 65 measures the probability that an individual was working at age $64\frac{1}{2}$ but was not at age $65\frac{1}{2}$. This panel looks very similar to the baseline case. In both cases job exit rates are near 20% at ages 62 and 65, and are lower at other ages. Furthermore, in both cases, job exit rates for those with retiree coverage are highest at age 62, whereas job exit rates for those with tied coverage are highest at age 65.

Because it seems extreme to treat an individual who leaves the labor force at age 65 and 1 day as exiting at age 66, we think measuring participation 6 months after a birthday yields more plausible results. Because measuring participation on survey dates gives similar results and drops fewer observations than measuring participation 6 months after a birthday, we use participation on survey dates as our measure of participation throughout.

Another measurement issue is the treatment of the self-employed. Our preferred approach is to include the self-employed in our analysis, and treat them as working with no health insurance. The lower right panel of Figure H.2 shows job exit rates when we drop the self-employed, but measure health insurance as in the baseline case. The main difference caused by dropping the

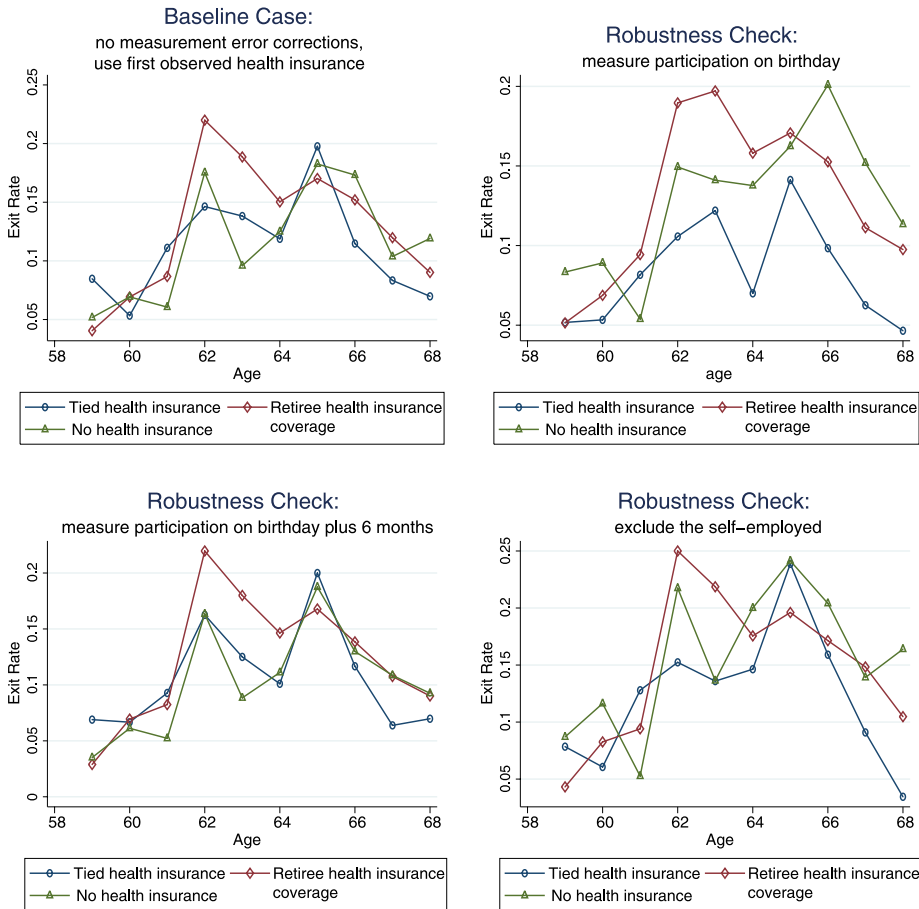


FIGURE H.2.—Job exit rates using different measures of labor force participation.

self-employed is that those with no health insurance have much higher job exit rates, especially at age 65. Nevertheless, those with retiree coverage are still most likely to exit at age 62 and those with tied and no health insurance are most likely to exit at age 65.

APPENDIX I: THE MEDICAL EXPENSE MODEL

Recall from equation (G.1) that health status, health insurance type, labor force participation, and age affect medical expenses through the mean shifter $m(\cdot)$ and the variance shifter $\sigma(\cdot)$. Health status enters $m(\cdot)$ and $\sigma(\cdot)$ through 0–1 indicators for bad health, and age enters through linear trends. On the other hand, the effects of Medicare eligibility, health insurance, and labor force

participation are almost completely unrestricted, in that we allow for an almost complete set of interactions between these variables. This implies that mean medical expenses are given by

$$m(H_t, I_t, t, P_t) = \gamma_0 H_t + \gamma_1 t + \sum_{h \in I} \sum_{P \in \{0,1\}} \sum_{a \in \{t < 65, t \geq 65\}} \gamma_{h,P,a}.$$

The one restriction we impose is that $\gamma_{\text{none},0,a} = \gamma_{\text{none},1,a}$ for both values of a , that is, participation does not affect health care costs if the individual does not have insurance. This implies that there are 10 $\gamma_{h,P,a}$ parameters, for a total of 12 parameters apiece in the $m(\cdot)$ and the $\sigma(\cdot)$ functions.

To estimate this model, we group the data into 10-year age (55–64, 65–74, 75–84) \times health status \times health insurance \times participation cells. For each of these 60 cells, we calculate both the mean and the 95th percentile of medical expenses. We estimate the model by finding the parameter values that best fit this 120-moment collection. One complication is that the medical expense model we estimate is an annual model, whereas our data are for medical expenses over 2-year intervals. To overcome this problem, we first simulate a panel of medical expense data at the 1-year frequency, using the dynamic parameters from French and Jones (2004) shown in Table III of the main paper and the empirical age distribution. We then aggregate the simulated data to the 2-year frequency; the means and 95th percentiles of this aggregated data are comparable to the means and 95th percentiles in the HRS. Our approach is similar to the one used by French and Jones (2004), who provided a detailed description.

Relative to other research on the cross-sectional distribution of medical expenses, we find higher medical expenses at the far right tail of the distribution. For example, Blau and Gilleskie (2006) used different data and methods to find average medical expenses that are comparable to our estimates. However, they found that medical expenses are less volatile than our estimates suggest. For example, they found that for households in good health and younger than 65, the maximum expense levels (which seem to be slightly less likely than 0.5% probability events) were \$69,260 for those without coverage, \$6,400 for those with retiree coverage, and \$6,400 for those with tied coverage. Table II in the main text shows that our estimates of the 99.5th percentile (i.e., the top 0.5 percentile of the distribution) of the distributions for healthy workers are \$86,900 for those with no coverage, \$32,700 for those with retiree coverage, and \$30,600 for those with tied coverage.

Berk and Monheit (2001) used data from the MEPS, which arguably has the highest quality medical expense data of all the surveys. Analyzing total billable expenses, which should be comparable to our data for the uninsured, Berk and Monheit found that those in the top 1% of the medical expense distribution have average medical expenses of \$57,900 (in 1998 dollars). Again, this is below our estimate of \$86,900 for the uninsured. This discrepancy is not surprising. Berk and Monheit's estimates are for all individuals in the population,

whereas our estimates are for older households (many of which include two individuals). Furthermore, Berk and Monheit's estimates exclude all nursing home expenses, while the HRS, although initially consisting only of noninstitutionalized households, captures the nursing home expenses these households incur in later waves.

APPENDIX J: THE PREFERENCE INDEX

We construct the preference index for each member of the sample using the wave-1 variables V3319, V5009, and V9063. All three variables are self-reported responses to questions about preferences for leisure and work. In V3319, respondents were asked if they agreed with the statement (if they were working), "Even if I didn't need the money, I would probably keep on working." In V5009, they were asked, "When you think about the time when you [and your (husband/wife/partner)] will (completely) retire, are you looking forward to it, are you uneasy about it, or what?" In V9063, they were asked (if they were working), "On a scale where 0 equals dislike a great deal, 10 equals enjoy a great deal, and 5 equals neither like nor dislike, how much do you enjoy your job?"

Because it is computationally intensive to estimate the parameters of the type probability equations in our method of simulated moments approach, we combine these three variables into a single index that is simpler to use. To construct this index, we regress labor force participation on current state variables (age, wages, assets, health, etc.), squares and interactions of these terms, the wave-1 variables V3319, V5009, and V9063, and indicators for whether these variables are missing. We then partition the $x\hat{\beta}$ matrix from this regression into $x_1\hat{\beta}_1$, where the x_1 matrix consists of V3319, V5009, V9063, and indicators for these variables being missing, and $x_2\hat{\beta}_2$, where the x_2 matrix contains all the other variables. Our preference index is $x_1\hat{\beta}_1$.

Individuals who were not working in 1992 were not asked any of the preference questions and are not included in the construction of our index. Because everyone who answered the preference questions worked in 1992, we estimate the regression models with participation data from 1998–2006.

Finally, we discretize the index into three values: out, for those not employed in 1992; low, for workers with an index in the bottom half of the distribution; and high for the remainder.

APPENDIX K: ADDITIONAL PARAMETER ESTIMATES

We assume that the probability of belonging to a particular type follows a multinomial logit function. Table K.I shows the coefficients of the preference type prediction equation. One interesting feature of this equation is that wealthy individuals who have no health insurance coverage have a high probability of being type-2 agents. Given that many of these individuals are entrepreneurs, it is not surprising that they are often placed in the "motivated" group.

TABLE K.I
PREFERENCE TYPE PREDICTION COEFFICIENTS

	Preference Type 1		Preference Type 2	
	Parameters	Std. Errors	Parameters	Std. Errors
Preference index = out	-5.33	1.04	-7.33	7.96
Preference index = low	4.79	1.99	0.18	1.32
Preference index = high	2.35	0.81	4.09	0.71
No insurance coverage	3.35	1.54	-2.45	1.82
Retiree coverage	-0.98	0.86	-0.32	0.46
Initial health ^a	-1.04	0.35	-0.37	0.27
Initial wages ^a	2.74	0.72	-1.01	0.44
Assets/wages ^a	-0.48	0.82	0.97	0.61
AIME/wages ^a	-0.25	0.70	-0.21	0.44
Health cost shock (ψ)	-1.16	0.65	0.22	0.47
Age - 60	-0.56	0.98	1.72	1.73
Assets ^a × (no ins. coverage)	-0.53	0.36	1.41	0.54

^aVariables are expressed as a fraction of average.

Table K.II shows the parameter estimates for the robustness checks. In the no-saving case, β and θ_B are both very weakly identified. We therefore follow Rust and Phelan (1997) and Blau and Gilleskie (2006, 2008) by fixing β , in this case to its baseline values of 0.95, 0.86, and 1.12 (for types 0, 1, and 2, respectively). Similarly, we fix θ_B to zero. Since the asset distribution is degenerate in this no-saving case, we no longer match asset quantiles or quantile-conditional participation rates, matching instead participation rates for each health insurance category.

The last column shows the parameter estimates that result when we remove the preference index described in Appendix J from our type prediction equations; we also remove the preference index-conditional moment conditions from the GMM criterion function. Although the coefficients of the type prediction equations change dramatically, the estimated preference parameters change very little.

REFERENCES

- ALTONJI, J., AND L. SEGAL (1996): "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353-366. [13]
- BERK, M., AND A. MONHEIT (2001): "The Concentration of Health Care Expenditures, Revisited," *Health Affairs*, 20, 9-18. [24]
- BLAU, D. (1994): "Labor Force Dynamics of Older Men," *Econometrica*, 62, 117-156. [22]
- BLAU, D., AND D. GILLESKIE (2006): "Health Insurance and Retirement of Married Couples," *Journal of Applied Econometrics*, 21, 935-953. [24,26]
- (2008): "The Role of Retiree Health Insurance in the Employment Behavior of Older Men," *International Economic Review*, 49, 475-514. [26]
- BOUND, J., T. STINEBRICKNER, AND T. WAIDMANN (2010): "Health, Economic Resources and the Work of Older Americans," *Journal of Econometrics*, 156, 106-129. [16]

TABLE K.II
ESTIMATED PREFERENCE PARAMETERS FOR THE ROBUSTNESS CHECKS^a

Parameter and Definition	Baseline	No Saving	Homogeneous Preferences	No Preference Index
γ : Consumption weight				
Type 0	0.412 (0.045)	0.302 (0.026)	NA	0.405 (0.040)
Type 1	0.649 (0.007)	0.583 (0.007)	0.550 (0.009)	0.647 (0.008)
Type 2	0.967 (0.203)	0.9999 (NA)	NA	0.986 (0.375)
β : Time discount factor				
Type 0	0.945 (0.074)	0.945 (NA)	NA	0.962 (0.064)
Type 1	0.859 (0.013)	0.859 (NA)	0.970 (0.009)	0.858 (0.014)
Type 2	1.124 (0.328)	1.124 (NA)	NA	1.143 (0.552)
ν : Coefficient of relative risk aversion, utility	7.49 (0.312)	6.35 (0.178)	5.78 (0.447)	7.61 (0.288)
θ_B : Bequest weight ^b	0.0223 (0.0012)	0.00 (NA)	0.0132 (0.0007)	0.0221 (0.0009)
κ : Bequest shifter, in thousands	444 (28.2)	0.00 (NA)	786 (26.9)	443 (21.3)
C_{\min} : Consumption floor	4,380 (167)	4,440 (152)	5,000 (167)	4,430 (211)
L : Leisure endowment, in hours	4,060 (44)	4,130 (70)	4,700 (63)	4,090 (44)
ϕ_H : Hours of leisure lost, bad health	506 (20.9)	939 (42.1)	303 (25.6)	509 (30.1)
ϕ_{P0} : Fixed cost of work: intercept, in hours	826 (20.0)	880 (24.4)	1,146 (36.0)	827 (28.8)
ϕ_{P1} : Fixed cost of work: (age – 60), in hours	54.7 (2.57)	36.5 (3.09)	16.9 (1.26)	52.7 (2.48)
ϕ_{RE} : Hours of leisure lost, reentering market	94.0 (8.64)	77.0 (12.4)	155.9 (13.6)	95.6 (24.5)
χ^2 Statistic	751	366	568	853
Degrees of freedom	171	86	169	145

^aA diagonal weighting matrix is used in the calculations. See Appendix F for details. Standard errors are given in parentheses.

^bThe parameter is expressed as the marginal propensity to consume out of final-period wealth.

- BUCHINSKY, M. (1998): "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *Journal of Human Resources*, 33, 88–126. [9,10]
- CHAMBERLAIN, G. (1992): "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business & Economic Statistics*, 10, 20–26. [10]
- CHERNOZHUKOV, V., AND C. HANSEN (2002): "An IV Model of Quantile Treatment Effects," Working Paper 02-06, MIT. [9]
- COMMITTEE ON WAYS AND MEANS, U.S. HOUSE OF REPRESENTATIVES (2000): *2000 Green Book*. Washington: United States Government Printing Office. [2,16]
- DAVID, M., R. LITTLE, M. SAMUHEL, AND R. TRIEST (1986): "Alternative Methods for CPS Income Imputation," *Journal of the American Statistical Association*, 81, 29–41. [17]
- DUFFIE, D., AND K. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952. [12]
- EPPLE, D., AND H. SEIG (1999): "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 107, 645–681. [10]
- FRENCH, E. (2005): "The Effects of Health, Wealth and Wages on Labor Supply and Retirement Behavior," *Review of Economic Studies*, 72, 395–427. [2,9]
- FRENCH, E., AND J. JONES (2004): "On the Distribution and Dynamics of Health Care Costs," *Journal of Applied Econometrics*, 19, 705–721. [11,21,24]
- GOURINCHAS, P., AND J. PARKER (2002): "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89. [9]
- GRUBER, J., AND B. MADRIAN (1996): "Health Insurance and Early Retirement: Evidence From the Availability of Continuation Coverage," in *Advances in the Economics of Aging*, ed. by D. A. Wise. Chicago: University of Chicago Press, 115–143. [20]
- GUSTMAN, A., O. MITCHELL, A. SAMWICK, AND T. STEINMEIER (2000): "Evaluating Pension Entitlements," in *Forecasting Retirement Needs and Retirement Wealth*, ed. by O. Mitchell, P. Hammond, and A. Rappaport. Chicago: University of Chicago Press, 309–326. [5]
- JUDD, K. (1998): *Numerical Methods in Economics*. Cambridge, MA: MIT Press. [8]
- LAIBSON, D., A. REPETTO, AND J. TOBACMAN (2007): "Estimating Discount Functions With Consumption Choices Over the Lifecycle," Working Paper, Harvard University. [9]
- LITTLE, R. (1988): "Missing Data Adjustments in Large Surveys," *Journal of Business & Economic Statistics*, 6, 287–301. [17]
- MANSKI, C. (1988): *Analog Estimation Methods in Econometrics*. New York: Chapman & Hall. [9]
- NEWBY, W. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256. [13]
- NEWBY, W., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing" in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: Elsevier. [13]
- PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057. [12,13]
- PISCHKE, J.-S. (1995): "Measurement Error and Earnings Dynamics: Some Estimates From the PSID Validation Study," *Journal of Business & Economic Statistics*, 13, 305–314. [13]
- POWELL, J. (1994): "Estimation of Semiparametric Models," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: Elsevier. [9,13]
- RUST, J., AND C. PHELAN (1997): "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, 65, 781–831. [26]
- SCHOLZ, J., A. SESHADRI, AND S. KHITATRAKUN (2006): "Are Americans Saving Optimally for Retirement?" *Journal of Political Economy*, 114, 607–643. [16]
- TAUCHEN, G. (1986): "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions," *Economics Letters*, 20, 177–181. [8]

Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604, U.S.A.; efrench@frbchi.org

and

*Dept. of Economics, University at Albany, SUNY, BA-110, Albany, NY 12222,
U.S.A.; jbjones@albany.edu.*

Manuscript received November, 2007; final revision received June, 2010.