How Social Security and Medicare affect retirement behavior in a world of incomplete markets

John Rust
Yale University
jrust@econ.yale.edu

Christopher Phelan
Northwestern University
cphelan@nwu.edu

Revised, April 1996[1]

Abstract: This paper provides an empirical analysis of how the U.S. Social Security and Medicare insurance system affect the labor supply of older males in the presence of incomplete markets for loans, annuities, and health insurance. We estimate a detailed dynamic programming (DP) model of the joint labor supply and Social Security acceptance decision, focusing on a sample of males in the low to middle income brackets whose only pension is Social Security. The DP model delivers a rich set of predictions about the dynamics of retirement behavior, and comparisons of actual vs. predicted behavior show that the DP model is able to account for wide variety of phenomena observed in the data, including the pronounced peaks in the distribution of retirement ages at 62 and 65 (the ages of early and normal eligibility for Social Security benefits, respectively). We identify a significant fraction of "health insurance constrained" individuals who have no form of retiree health insurance other than Medicare, and who can only obtain fairly priced private health insurance via their employer's group health plan. The combination of significant individual risk aversion and a long tailed (Pareto) distribution of health care expenditures implies that there is a significant "security value" for these individuals to remain employed until they are eligible for Medicare coverage at age 65. Overall, our model suggests that a number of heretofore puzzling aspects of retirement behavior can be viewed as artifacts of particular details of the Social Security rules, whose incentive effects are especially strong for lower income individuals and those who do not have access to fairly priced loans, annuities, and health insurance.

[1] Forthcoming, *Econometrica*. This paper is an abridged version of a monograph prepared for the NAKE workshop on "Pensions and Public Finance in an Aging Society" at the Tinbergen Institute, Rotterdam in December 1993. We thank the workshop participants and the organizers Lans Bovenberg and Caspar van Ewijk for their feedback. We received additional helpful comments from participants of the 1994 conference on "The Future of the Welfare State" sponsored by the *Scandinavian Journal of Economics* and from participants of seminars and workshops at several universities and organizations in the U.S. and Europe. We are especially grateful for extremely helpful and detailed comments from D. Blau, G. Burtless, H.P. van Dalen, A. Gustman, M. Hurd, R. Moffitt, J. Smith, B. Sørensen, T.L. Steinmeier, S. Stern, D. Wise and three anonymous referees. Various stages of this research have been supported by grants from the U.S. National Institute of Aging, the National Science Foundation, the Bradley Foundation, the Institute for Empirical Macroeconomics at the Federal Reserve Bank of Minneapolis, the Graduate School of the University of Wisconsin, and the LaFollette School of Public Policy at the University of Wisconsin. Needless to say, none of the above agencies or individuals bear any responsibility for the views expressed within this paper.

1. Introduction

This paper provides an empirical analysis of how the U.S. Social Security and Medicare insurance system affects the labor supply of older workers, with explicit recognition of the fact that certain individuals do not have access to fairly priced loans, annuities, and health insurance. We estimate a dynamic programming (DP) of an individual's decision-making process about how much to work and when to apply for Social Security benefits using the Retirement History Survey (RHS), a comprehensive panel data set following a sample of individuals initially aged 58-63 from 1969 to 1979 in 6 biennial interviews. We focus on a subsample of low to middle income males whose only "pension" is Social Security. The DP model incorporates constraints imposed by incomplete markets and embeds the rules governing payment of Social Security old age and Medicare benefits into individuals' expectations about future income streams. We find that careful modeling of both of these aspects is essential to understanding observed retirement behavior. The most prominent feature of the data is that most men exit the labor force in their mid-60's, with pronounced peaks in the retirements at ages 62 and 65 — precisely the ages of eligibility for early and normal Social Security benefits, respectively. The DP model "explains" the peak at age 62 as a result of borrowing constraints (it is illegal to borrow against one's future Social Security benefits), and the fact that the men in our sample do not have private pensions and few have accumulated sufficient tangible net worth to be able to finance significant retirement consumption prior to age 62.[1] The DP model "explains" the peak in retirements at age 65 as a result of an incomplete annuities market and the fact that the Social Security benefit formula is actuarially unfair for retirements after age 65. Although private annuity contracts certainly exist, very few people in our sample ever purchased them, possibly because most annuities offer notoriously poor returns as Friedman and Warshawsky (1988) have documented. For all practical purposes the only retirement "annuity" available for our sample of men is Social Security.

However we find that the actuarial unfairness of Social Security is able to account for only part of the peak in retirements at age 65. The remainder of the age 65 peak is explained by another form of market incompleteness — incomplete health insurance — and the fact that Medicare insurance is available only to individuals over 65 who have applied for Social Security benefits.[2] Although the overall effect of Social Security is to create strong disincentives to continued labor force participation, it creates strong incentives for certain individuals to remain employed up until their 65^{th} birthday. We identify a significant fraction of "health insurance constrained" individuals who have no form of retiree health insurance other than Medicare, and who can only obtain fairly priced private health insurance via their employer's group health plan. The combination of significant individual risk aversion and a long tailed (Pareto) distribution of health care expenditures implies that there is a significant "security value" for these individuals to remain employed until they become eligible for Medicare coverage at age 65. Overall, our findings suggest that a

1

2

number of previously puzzling aspects of retirement behavior can be viewed as artifacts of particular details of the Social Security rules, whose incentive effects can be especially strong for lower income individuals and those who do not have access to fairly priced loans, annu ities, and health insurance.

This is not the first paper to examine these issues. We characterize the existing literature as presenting three puzzles that we attempt to address in this paper: 1) the *early retirement puzzle,* 2) the *age 65 retirement puzzle,* and 3) the *Medicare puzzle.* A key question underlying all of these puzzles is whether Social Security policy has an important influence on individual behavior. Although there is little dispute about the basic facts concerning the historic decline in older male labor force participation (which began shortly after the introduction of Social Security in 1935, see Ransom and Sutch 1988), "proving" that Social Security is the major cause of this decline turns out to be a surprisingly difficult task. Econometric studies of this issue have generated very mixed conclusions: studies such as Boskin (1977), Boskin and Hurd (1978), Burkhauser (1980), Parsons (1982), Gustman and Steinmeier (1986), Kahn (1988), and Stewart (1995) have found that Social Security (and the closely related disability insurance program) have had strong negative effects on male labor supply, whereas other studies such as Blinder, Gordon and Wise (1980), Fields and Mitchell (1984), Moffitt (1987), Burtless (1986), and Krueger and Pischke (1991) have concluded that the large increases in real Social Security benefits over the past four decades have had little effect on behavior. These latter studies have attributed the large apparent effects of Social Security to a variety of other factors: Lumsdaine and Wise (1994) attributed the decline in labor force participation rates of older males to the rise in private pensions; Burtless (1986) suggested factors such as "rising personal wealth levels, sharply higher unemployment levels in the period after 1970, and changing attitudes toward work and retirement." (p. 801), and Krueger and Pischke (1991) speculated that the decline may be due to "a possible reduction in the average health of the elderly" (p. 25).

There is also disagreement about whether the large peaks in retirements at 65 and 62 are a result of Social Security policy. The "option value" and dynamic programming models of Stock and Wise (1991) and Lumsdaine, Stock and Wise (1992, 1994, 1995) systematically underpredict the peak in retirements at age 65.[3] After investigating several alternative economic explanations for the peak in retirements at age 65, Lumsdaine, Stock and Wise (1995) arrived at the conclusion that it is a result of "social custom and the use of an age 65 rule of thumb" (p. 19). On the other hand the dynamic structural model of Gustman and Steinmeier (1986) does a very good job of capturing the peaks in retirements at ages 62 and 65. To our knowledge the Gustman-Steinmeier model is the first study to show that these peaks can be explained entirely on economic grounds. Indeed "When the effects of pensions, Social Security, and mandatory retirement are eliminated from the compensation function, the peaks in retirement at ages 62 and 65 disappear *completely.*" (p. 580). However, it is not entirely clear why the Gustman-Steinmeier model is successful in capturing the peaks in retirements at these ages, since their model abstracts from uncertainty and assumes that

[3]

individuals have perfect ability to borrow against future Social Security benefits. Steinmeier (private communication) suggested that the peak at age 62 is due to private pensions: "Although we did not test this explicitly, I am fairly confident that the peak at age 62 reflects the drop in compensation attributable to the lower pension accruals after that age." However this reasoning cannot explain the fact that the peak in retirements at age 62 is equally large for our sample of men *without* private pensions.

Finally several recent studies have come to opposite conclusions about the impact of Medicare and retiree health insurance on labor supply behavior. Recent dynamic structural models have been unable to uncover a strong link between health insurance and labor supply behavior. For example Lumsdaine, Stock and Wise (1994) concluded that "indirect evidence suggests that Medicare eligibility is not an important determinant of the age-65 retirement effect" (p. 19), and Gustman and Steinmeier (1993) concluded that "The omission of retiree health benefits from the opportunity set in most previous studies of retirement behavior is not likely to invalidate the conclusions of these studies, with regard to the effects on retirement of pensions, social security, or other components of the opportunity set" (p. 32). These conclusions may be an artifact of an overly simplistic treatment of health care risks. Both studies value Medicare at the *expected value* of per capita reimbursements and treat this as an addition to the person's monthly Social Security benefit. However if individuals are risk averse and the distribution of health care costs is highly skewed, the person's *certainty equivalent* valuation of Medicare benefits will be far larger than the expected value of per capita reimbursements. This may be why recent reduced-form studies have found that health insurance coverage does have very strong effects on retirement behavior. For example, Gruber and Madrian (1993) found that even the limited term "continuation coverage" mandates "have a sizable and significant effect on retirement" (p. 29).

Section 2 reviews some of the key facts about retirement behavior from the RHS survey in order to motivate our particular specification of the DP model. An important finding of this section is that unhealthy individuals are more than twice as likely as healthy individuals to apply for Social Security benefits at age 62. Further, we identify a subpopulation of "health insurance constrained" individuals who have no form of retiree health insurance other than Medicare and who can only obtain fairly priced private health insurance via their employer's group health plan. We show that individuals in this group are nearly four times more likely than unconstrained individuals to apply for Social Security benefits at the normal retirement age of 65 (when they become eligible for Medicare) than at the early retirement age 62. Section 3 presents the DP model which is capable of explaining these facts. Our formulation of the DP model treats labor supply and application for Social Security benefits as separate decisions, a distinction that has been ignored in most previous studies. Section 4 summarizes the estimation results for individuals' beliefs about future mortality, health status, marital status, earnings, and health expenditures, which we model as a collection of age, state, and decision-dependent conditional probability distributions. We find that incorporation of subjective uncertainty about health status and health care expenditures is particularly important for understanding the retirement process. Section 5 presents estimates of individuals' preferences for leisure and consumption and section 6 provides a detailed analysis of the model's predictions and its ability to fit the data. Our two main conclusions are: 1) a DP model with a relatively parsimonius specification of preferences is able to provide a surprisingly good representation of the complex dynamics of retirement decision making for our sample of RHS males, 2) the DP model is able to do this because we have accurately modeled the nature of the "incentive schemes" created by the interaction of the U.S. Social Security rules and the pattern of incomplete private markets facing individuals. Although individuals have "smooth" preferences in our model (i.e. their indifference curves for leisure versus consumption do not change rapidly at any particular age or age range), the Social Security incentives are quite discontinuous, especially at ages 62 and 65. Thus, our DP model provides a simple explanation for the observed discontinuities in behavior, particularly the peaks in retirements at ages 62 and 65: it is a "best response" by rational individuals to incentives created by the Social Security rules, the constraints imposed by legal restrictions, and incomplete markets. Section 7 summarizes our conclusions and discusses some limitations of our model.

2. Social Security, Medicare, and Labor Force Participation of Older Males: Some Key Facts

Before presenting our relatively detailed structural model of retirement behavior, it is useful to provide a summary of several key facts about retirement behavior that have emerged from empirical analysis of the RHS data (see, e.g. Blau, 1994, Ruhm, 1990, and Rust, 1989, 1990) that motivated our particular specification of the DP model presented in section 3:

1. there is a great deal of heterogeneity in individual labor supply paths, and "The prevalence of labor force transitions at older ages appears ... to have been significantly underestimated in previous studies. The quarterly data also reveal a strikingly high peak of close to 25% in the rate of labor force exit at the exact age 65." (Blau, 1994, p. 119).

2. the "standard model" which treats retirement as an absorbing state is incapable of explaining the labor supply behavior of the majority of the RHS sample: "Fewer than two-fifths of household heads retire directly from career jobs, over half partially retire at some point in their working lives, and a quarter reenter the labor force after initially retiring." (Ruhm, 1990, p. 482),

3. although aggregate labor supply profiles look rather smooth (suggesting that the "representative agent" makes a gradual transition from full time work into retirement), individual labor supply trajectories are anything but smooth with many individuals making discontinuous transitions from full time work to not working (Rust, 1990).

4. relatively few individuals gradually phase out of the labor force by reducing hours of work on their main career jobs: "Instead, the majority of individuals leave career jobs well before retirement and enter a transitional job-stopping period consisting of some combination of bridge employment, partial retirement, and reverse retirement." (Ruhm, 1990, p. 497),

5. at any point in time there is a great deal of clumping in the cross-sectional distributions of hours of work with most individuals working either 0 and 2000 hours per year. The peak at 2000 corresponds to full-time work (40 hours/week $\times$ 50 weeks/year). The relatively small mass of inidividuals with part-time or part-year jobs are approximately uniformly distributed over the $(0, 2000)$ interval (Rust, 1990).

The finding that among the of their career jobs, and those who do gradually reduce hours of work typically do so by taking on a sequence of lower wage partial retirement "bridge jobs" rather than gradually reducing hours of work at their full-time pre-retirement "career job" suggests the existence of explicit or implicit constraints on the individual's choice of hours of work.[4] This suggests that it is inappropriate to treat hours of work as a continuous choice variable. Furthermore, the finding that most of the mass in the distribution of annual hours of work is at 0 and 2000 implies that we would not lose a great deal of information by discretizing annual hours of work into three cells:

$$F = \text{annual hours} > 1600$$

$$P = \text{annual hours} \in (300, 1600]$$

$$R = \text{annual hours} \leq 300.$$

We adopt this discretization in the subsequent analysis, departing from the predominant Euler equation approach to the analysis of dynamic labor supply decisions which assumes that hours of work is a continuous choice variable (see, e.g. MaCurdy, 1983, 1985).[5]

It should come as no surprise that for most individuals, the labor supply decision is closely coordinated with the decision of whether or not to begin collecting Social Security benefits. Figure 2.1 provides a key insight into this relationship by plotting the distributions of ages of first entitlement and first receipt of Social Security benefits. The two distributions differ because the decision to apply for Social Security benefits is logically distinct from the labor supply decision. Note that an individual is defined to be first entitled to Social Security benefits once they are age 62 and have accumulated 40 quarters of coverage, *and have filed an application for benefits.* However the age at which the individual first *receives* Social Security benefits will be later than the age of first entitlement if the individual's Social Security benefits are taxed away due to the "earnings test". The earnings test will be described in more detail in section 4, but is basically a 50% tax on wages earnings in excess of a small threshold or "test" level. From figure 2.1 it is apparent that while most individuals begin receiving Social Security benefits in the same year that they apply, there is a significant fraction of individuals who apply for Social Security benefits prior to 65 but don't begin receiving benefits until their late 60's or early 70's due to continued employment at full time jobs. Many of the individuals in this latter category are professionals, self-employed, and other higher income individuals who do not appear to have any intention to "retire" (at least by the normal retirement age of 65). Any reasonable theory needs to explain why these individuals would apply for Social Security benefits at age 65 or earlier even though they have evidently planned to continue working well into their 70's. At the same time, the theory also needs to explain why the labor force exit and Social Security application decisions are so closely coordinated for other individuals, particularly at ages 62 and 65.

---

[4]

[5] MaCurdy notes that even in cases where hours of work can be plausibly treated as a continuous choice variable, the Euler equation approach runs into difficulties when applied to retirement behavior due to the prevalence of binding corner solutions at zero hours of work: "Because the above procedures ignore statistical problems relating to the endogeneity of labor force participation decisions, they are of limited use in estimating period specific utilities associated with households in which corner solutions for hours of work are not a certainty for at least one family member which, of course, includes households with wives and older households where retirement may occur." MaCurdy, (1983), p. 276–277. However as we will see in sections 4 and 5, our simple 3 state discretization of hours of work may be too coarse: we would advise the use of finer discretizations in future work.

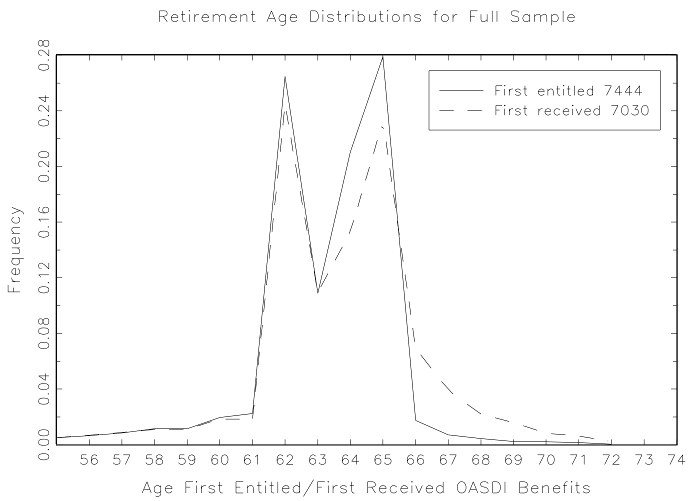

Figure 2.1 Distributions of Ages of Application and First Receipt of Social Security Benefits

Figures 2.2 and 2.3 present two final insights into retirement behavior from the RHS data. Figure 2.2 plots the distribution of retirement ages for two groups of individuals:[6] 1) those who are in bad health (i.e. who reported having a health problem that limited their ability to work or get around in the majority of the survey interviews), and 2) the remaining sample (who we classified as being in good health by default). The figure provides strong evidence of self-selection of retirement benefits by health status: people in bad health are nearly twice as likely to apply for Social Security at 62 than 65 whereas the distribution for those in good health is nearly the mirror image. Figure 2.2 suggests that many of the individuals who are in bad health may not have been in sufficiently bad health to qualify for DI benefits, and it is reasonable to suppose that many of these individuals would have liked to begin receiving Social Security benefits prior to 62 if it were possible. If this is true, then a change in the age of eligibility for early retirement benefits ought to have a large impact on the labor supply and welfare of unhealthy individuals.

Figure 2.3 plots the distributions of retirement ages for four different groups of individuals with different types of health insurance: *eph* are individuals who we determined to have employer provided private health insurance but no access to retiree health insurance, *gph* are individuals who have private health plans such as Blue Cross/Blue Shield that provide coverage independent of whether or not they are employed, *mca* are individuals who have qualified for Medicaid insurance, and *nhi* are individuals who don't have any type of health insurance. [7] The difference in the

[6] The figure excludes individuals who qualified for Social Security DI benefits.

[7] The basis on which we classified individuals to each of these groups will be described in more detail in section 4 and appendix 1.

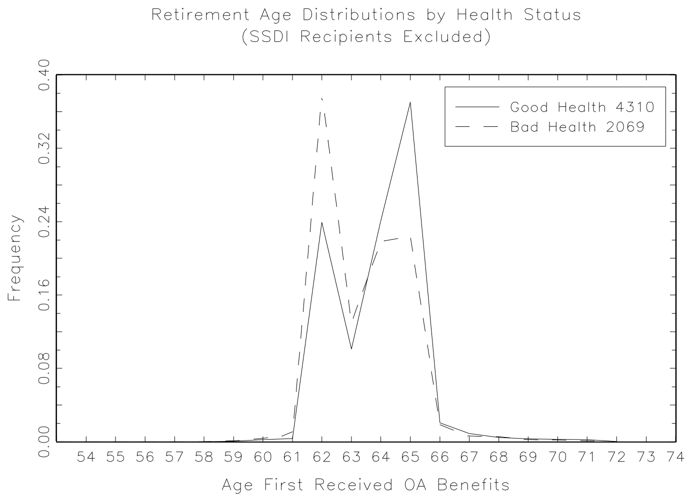

Figure 2.2 Impact of Health Status on the Age of Application for Social Security Benefits

distribution of application ages for the *eph* group is striking.[8] Overall, figure 2.3 suggests that Medicare insurance (which is available to those who are over 65 and eligible for Social Security benefits) could have big effects on the retirement plans of certain groups of individuals, particularly those who do not have access to retiree health insurance. These "health insurance constrained" individuals in the *eph* group are nearly twice as likely to apply for Social Security at age 65 than at age 62. The other groups have no such incentive since they either have no health insurance to begin with (*nhi*), or have health insurance that isn't linked to employment (*mca* and *gph*). Figure 2.3 shows that individuals in the latter groups are up to 4 times more likely than the *eph* group to apply for Social Security benefits at age 62 than age 65. We conjecture that risk aversion about uninsured health care expenditures could create strong incentives for individuals in the *eph* group to remained employed (where they are covered by their employer's health plan) until they turn 65 and are eligible for Medicare. Taken as a whole, the results of this section constitute strong *prima facie* evidence of the need for an integrated model of the interaction of Social Security, Medicare, and incomplete markets in order to understand the dramatic variations in retirement behavior of different groups of individuals.

[8] A plot of the distribution of age of first receipt of Social Security benefits is very similar to the distribution of age of application presented in figure 2.3. The main difference is a reduction in the peak at age 65 for the *eph* group: although nearly 40% of this group applies for benefits at age 65, only 30% actually first start receiving benefits at age 65 and the remaining 10% continue working at their full-time jobs until their late 60's or early 70's, losing all their benefits to the earnings test.

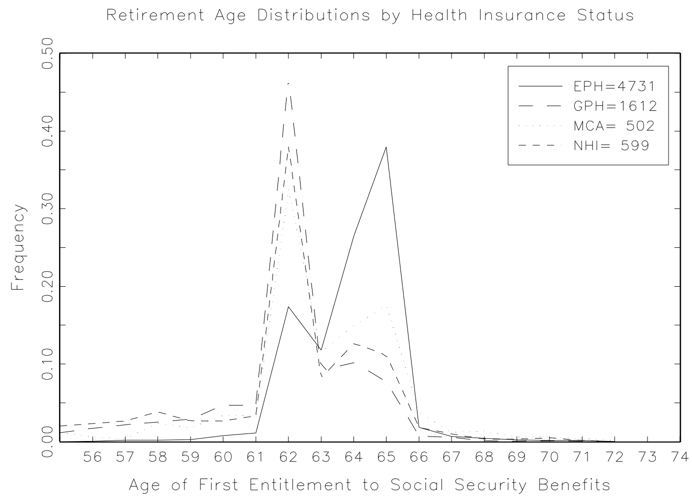

Figure 2.3 Impact of Health Insurance on the Age of Application and First Receipt of Social Security Benefits

3. A DP Model of the Impact Social Security and Medicare on Retirement Behavior

The main motivation for our approach, articulated in more detail in Phelan and Rust (1993), is that the best way to understand the welfare and behavioral impacts of Social Security, Medicare, and private pension and health insurance plans is to view them abstractly as *dynamic incentive schemes:* namely, as sets of rules governing a state and decision-dependent stream of future payoffs, with state-dependence arising from certain potentially insurable contingencies (such as sickness and death), and decision-dependence arising from the individual's choice of what age to apply for benefits and their subsequent labor supply behavior. This not only requires modeling of the nonlinearities in the "Social Security budget set" at any point of time (as was done in the pioneering work of Burtless and Mofitt, 1984), it also requires us to specify how current decisions affect future budget sets.

Dynamic programming (DP) provides a framework that is rich enough to accurately model the dynamic structure of the Social Security rules and the uncertainties and sequential nature of individuals' decision making processes. The DP framework imposes a very clear but structured view of the data, dichotomizing it into *state variables* s_t and *control variables* d_t. The basic structure of the DP problem is extremely simple. The decision maker's preferences and beliefs are specified by three objects: an intertemporal discount factor β, a single period utility function over states and decisions, $u_t(s, d, \theta_u)$, and a transition probability density $p_t(s_{t+1}|s_t, d_t, \theta_p, \alpha)$, representing the individual's subjective beliefs about uncertain future events. The "behavior" implied by the DP model is embodied by an *optimal decision rule* $\delta = (\delta_0, \ldots, \delta_T)$ where $d_t = \delta_t(s_t)$ specifies an age t individual's optimal decision d_t as a function of the realized state s_t, where for concreteness we have assumed that the individual dies with probability 1 at the terminal age T.[9] The sequence of decision rules δ is chosen to maximize the individual's expected discounted utility

$$V_t(s) = \max_{\delta} E_\delta \left\{ \sum_{j=t}^{T} \beta^{j-t} u_t(s_j, d_j, \theta_u) \Big| s_t = s \right\} \tag{3.1}$$

where $V(s_t)$ is the *value function* representing the expected discounted utility of an individual who is in state s_t and follows an optimal policy from time t onwards. The value function V and associated decision rule δ depend on the underlying primitives (u_t, p_t), $t = 1, \ldots, T$, which depend in turn on two sets of parameters: 1) a vector α codifying the details of Social Security and Medicare policy, and 2) a vector $\theta = (\beta, \theta_u, \theta_p)$ characterizing individuals' preferences and beliefs.

The α vector includes such components as the age of normal retirement, the benefit reduction factor for early retirement, the bend points and slopes describing the nonlinear relationship between average monthly wage and the individual's primary insurance amount (PIA), the level of Medicare deductible and coinsurance rates, and so forth. The details of exactly how the Social Security rules are encoded as a vector α and embedded in $p_t(s_{t+1}|s_t, d_t, \theta_p, \alpha)$ will be described in section 4.

Individuals are assumed to know their own "true" value of θ, call it θ^* (since they know their own preferences and beliefs), but θ^* is unknown from the standpoint of the econometrician who must infer it from the individual's observed sequence of states and decisions. We now impose some additional structure on the DP problem to enable us to do statistical inference using observations $\{s_t^i, d_t^i\}$ on the realized states and decisions of a panel of individuals $i = 1, \ldots, I$ where each individual is followed for $t = 1, \ldots, T_i$ time periods. We adopt the framework of Rust (1987,1988,1994,1995) and partition the state variable into two components, $s = (x, \epsilon)$ where x is a vector of state variables observed by both the econometrician and the individual, and ϵ is a vector observed only by the individual. We assume that the vector ϵ has as many components as the number of feasible actions in the individual's state-dependent choice set, $D_t(x)$, which specifies the set of feasible actions available to the individual in state x at time t. Thus $\epsilon_t(d)$ should be thought of as the net utility or disutility to taking action d at time t due to factors that are unobserved by the econometrician. Even though the DP problem delivers an optimal decision rule $d_t = \delta_t(x_t, \epsilon_t, \theta, \alpha)$ that is a deterministic relation from the standpoint of an individual who observes both (x_t, ϵ_t), it is random from the standpoint of an econometrician who only observes x_t. By integrating out the unobserved state variables from the optimal decision rule, we obtain a *conditional choice probability* $P_t(d|x, \theta, \alpha)$ that provides a basis for estimating the unknown parameter vector θ^* and simulating the impacts of alternative Social Security policies α:

$$P_t(d|x, \theta, \alpha) = \int I\{d = \delta_t(x, \epsilon, \theta, \alpha)\} q(d\epsilon|x). \tag{3.2}$$

---

[9] Of course, there is also a positive probability that the individual will die before the terminal period T. In this paper we assume that $T = 102$.

If we assume that $q(\epsilon|x)$ is a multivariate extreme value distribution, then Rust (1987, 1988) showed that $P_t(d|x,\theta,\alpha)$ has a multinomial logit representation

$$P_t(d|x,\theta,\alpha) = \frac{exp\{v_t(x,d,\theta,\alpha)\}}{\sum_{d' \in D(x)} exp\{v_t(x,d',\theta,\alpha)\}}, \tag{3.3}$$

where $D(x)$ denotes the individual's choice set in state x and v_t is the *expected value function* defined recursively by

$$v_t(x_t,d_t,\theta,\alpha) = u_t(x_t,d_t,\theta_u) + \beta \int log\Big[\sum_{d_{t+1} \in D(x_{t+1})} exp\{v_{t+1}(x_{t+1},d_{t+1},\theta,\alpha)\}\Big] p_t(dx_{t+1}|x_t,d_t,\theta_p,\alpha). \tag{3.4}$$

Solving (3.4) is equivalent to solving the DP problem by backward induction from the terminal age T. The expected value function $v_t(x,d,\theta,\alpha)$ is related to the value function $V_t(s_t) = V_t(x,\epsilon)$ defined in equation (3.1) by the identity

$$V_t(x,\epsilon) = \max_{d \in D(x)} [v_t(x,d,\theta,\alpha) + \epsilon(d)]. \tag{3.5}$$

The assumption that unobservable state variables follow an *IID* extreme value process can be criticized as an unrealistic and unjustified parametric restriction. However given that unobservables are deeply embedded into the solution of the DP problem, it is very difficult to incorporate more general forms of serial correlation in $\{\epsilon_t\}$ or depart substantially from the extreme value specification for the marginal distributions of ϵ_t and still obtain a computationally tractable econometric model.[10] If there is any virtue in the necessity of the *IID* assumption, it's the discipline it imposes on the modeling exercise: the only way the DP model can capture serially correlated retirement dynamics is via dependence in the observed state variables, with more careful attention to the economic rather than the statistical specification of the model. However we are not entirely free to impose arbitrary patterns of dependence in the observed state variables in order to get the choice probabilities (3.3) to fit the data since we will be imposing the additional hypothesis of *rational expectations.* This implies that the serial dependence in the $\{x_t\}$ process will be estimated (and therefore determined) by the data.[11]

Given panel data $\{x_t^i, d_t^i\}$, $t = 1,\ldots,T_i$, $i = 1,\ldots,I$ on the observed states and decisions of I individuals under some fixed Social Security policy, say α_{72} (corresponding to the Social Security law in effect after the 1972

[10] Some authors such as Berkovec and Stern (1991) have been successful in estimating DP models with a factor structure for unobservables $\epsilon_t = \eta + \nu_t$ where η is a person-specific random effect and ν_t is an *IID* process. Estimation of models that allow more flexible patterns of serial dependence is much harder, although solution algorithms based on monte carlo methods such as Keane and Wolpin, (1994) and Rust, (1995) offer the hope that models with more realistic patterns of serial dependence will be estimable in the near future.

[11] Even if we did have the luxury of fitting the DP model by imposing flexible forms of serial dependence in $\{\epsilon_t\}$, it is not clear whether the resulting model would yield an entirely satisfactory "explanation" of observed retirement behavior. In particular, one might suspect that serial correlation in $\{\epsilon_t\}$ reflects an incomplete specification of both the economic environment and the observable components of the DP model, raising difficult questions as to whether (and how) the $\{\epsilon_t\}$ process might be affected by changes in the Social Security policy parameters α. The approach of this paper is to start with a fairly comprehensive specification of the observable state vector x_t, including the most of the key state variables affecting the retirement decision process so that it is more plausible that the remaining unobserved state variables can be approximated as *IID* "noise" whose distribution is invariant to changes in α.

Social Security amendments), one can estimate individuals' preference parameters by finding the value $\hat{\theta}$ such that the predictions of the DP model "best fits" the data. In our case, the value of $\hat{\theta}$ that best fits the data is defined as the parameter value that maximizes the likelihood function $L(\theta)$ defined by:

$$L(\theta) = L(\beta, \theta_u \theta_p) = \prod_{i=1}^{I} \prod_{t=1}^{T_i} P_t(d_t^i | x_t^i, \theta, \alpha_{72}) p_t(x_t^i | x_{t-1}^i, d_{t-1}^i, \theta_p, \alpha_{72}), \tag{3.6}$$

As we will see in sections 4 and 5, while a relatively small number of parameters are used to specify individuals' preferences, $u_t(x_t, d_t, \theta_u)$, a very large number of parameters are needed to specify their beliefs, $p_t(x_t|x_{t-1}, d_{t-1}, \theta_p, \alpha_{72})$. Estimation is feasible only using a simpler two-stage estimation procedure described in Rust (1987, 1988, 1995) and summarized briefly here: 1) the θ_p parameters are estimated using a first stage partial likelihood function involving only products of the p_t terms, 2) using the initial consistent estimates $\hat{\theta}_p$ from the first stage, we solve the DP recursion (3.4) numerically and estimate the remaining parameters (β, θ_u) using a second stage partial likelihood function consisting of products of the choice probabilities P_t. As is well known (see Rust, 1988, 1995), the two stage procedure is generally not as efficient as full maximum likelihood estimation using the full likelihood function (3.6), and estimation error from the first stage parameter estimates $\hat{\theta}_p$ "contaminates" the estimated covariance matrix for θ_u in the second stage. Although Rust (1995) provides formulae for the "corrected" second stage covariance matrix, the computational burden required to carry out these corrections is as large as the burden involved in computing the full maximum likelihood estimates. Neither are computationally practical at this point.[12]

We now turn to the specification of the observed state and control variables, (x_t, d_t). In order to estimate a "realistic" specification of the DP model, we need fairly detailed information on the relevant states and decisions of individuals. In particular we need panel data on individuals' earnings histories and Social Security application decisions, and comprehensive information on their health, marital status, and employment status. There are few data sets that contain all the required information. The Retirement History Survey (RHS) – a panel study that interviewed over 11,000 households born between 1906 and 1911 biennially between 1969 and 1979 – is the only currently available data set that has sufficient breadth and depth to permit estimation of a DP model that adequately approximates the individual's decision making process.[13] In addition to offering a comprehensive source of data on health status, health and life insurance coverage, income, wealth, labor supply, consumption and health care expenditures, the RHS also has a linked set of Social Security earnings histories (SSER) and benefit records (SSMBR) that enable detailed modeling of the exact timing and levels of individual applications, payments, and terminations of Social Security benefits.

---

[12] If the information matrix corresponding to the partition $\theta = (\theta_u, \theta_p)$ is block-diagonal, then the two stage estimation procedure is asymptotically equivalent to full information maximum likelihood and estimation error in the first stage parameter estimates $\hat{\theta}_p$ does not contaminate the second stage covariance matrix estimate for θ_u. Although we have no way of knowing whether block-diagonality holds in this case, we have found that in a variety of other problems that the full information matrix is typically approximately block diagonal.

[13] The RHS will be superseded by the forthcoming Health and Retirement Survey (HRS). Unfortunately only the first wave of the HRS was available at the time this paper was written, and it is impossible to estimate beliefs (transition probabilities) using only one wave of survey data.

Our study focuses on male heads of households since a large fraction of female heads of households in the RHS are widows who we believe have substantially different labor force histories and retirement behavior than males. The other major sample restriction is to exclude over 44% of the sample (3,593 out of the full sample of 8,131 male head of households) who expected to receive benefits from private pension plans. We excluded these individuals due to the fact that the RHS has very sketchy information on the details of private pension plans. Similar to Social Security, it is critical to model the important details of the various pension plan provisions in order to understand their strong behavioral impacts. Rather than contaminate the DP model with crude "guess-estimates" of the provisions of typical pension plans, we decided to focus on Social Security whose provisions are known and therefore can be accurately modeled. The version of the DP model presented in this paper also excludes approximately 10% of the male population who qualified for Social Security DI benefits at any point during the decade of RHS survey. We excluded these individuals due to the fact that the SSMBR data only identify individuals who qualified for DI benefits, but does not identify those who applied and were rejected. In order to accurately model the DI application decision, we need to have estimates of the probability of being denied benefits conditional on health status, career history, age, and other factors.

After accounting for mortality, attrition, and various other exclusions we obtained an estimation subsample of 2,599 men, or a total of 7,574 person-year observations.[14] Since we will be focusing on transitions over two year intervals, we record employment status in the even years preceding the survey dates (which occurred on odd years). Thus, an individual's employment state at the 1969 RHS interview is determined by hours of work in 1968, and the employment decision at the 1969 interview is treated as a precommittment to a plan that takes effect two years later in 1970. Although it is theoretically possible to estimate DP models formulated at much finer time scales and "integrate out" the unobserved intervening states and decisions, as is well known there is a "curse of dimensionality" that quickly makes solution and estimation of very fine-grained DP models intractable. This paper can be viewed as an exploration into what can be accomplished with a relatively coarse 2-year discretization.

The RHS has sufficient detail to formulate a DP model with a 7-dimensional vector of observed state variables $x_t = (y_t, ss_t, e_t, m_t, h_t, hi_t, aw_t)$ and a 2-dimensional vector of control variables $d_t = (ed_t, ssd_t)$ defined by:

State variables:

$y_t =$ total family income (net of out of pocket expenditures on health care) in 1968 \$ discretized into *ys* intervals. In our specification $ys = 25$ with income ranging from $[-4000, 20000)$ in \$1,000 increments.[15] Total income

[14] The sample exclusions included elimination of observations due to missing data, internally inconsistent responses, and individuals with "too much" income or net worth (in excess of \$200,000 and \$1,000,000 1968 dollars, respectively). For a complete description of the exclusions, the Gauss program `sample.gpr` and all other data and estimation software is available via the Web site `http://thor.econ.wisc.edu`

[15] Since it is possible that out of pocket health expenditures could exceed income in any given year, we allow the support of the income distribution to have negative values.

is defined as the sum $y_t = rw_t + sw_t + rss_t + sss_t + oi_t - hc_t$ where rw_t is respondent wage income, sw_t is spouse wage income, rss_t is respondent Social Security benefits, sss_t is spouse Social Security benefits, oi_t is other (asset) income, and hc_t are out of pocket expenses on health care (net of insurance reimbursements).

e_t = employment state of the individual at age t discretized into *es* intervals. In our biennial model we set $es = 3$ with $e_t \in \{[0, 300), [300, 1600], (1600, \infty)\}$ being interpreted as the states NE (not employed), PT (part-time) and FT (full-time work) determined from hours worked in the year preceding time t.

ss_t = Social Security state of the individual at age t. In our biennial specification ss_t has 3 possible values: $ss_t \in \{\text{NE}, \text{ER}, \text{NR}\}$ where NE denotes that the individual is not eligible for Social Security benefits (under age 62 or hasn't yet applied), ER denotes an early retiree (first eligible when aged 62-64), and NR denote an normal retiree (first eligible after age 64).

m_t = marital status at age t, $\{1 = \text{single}, 2 = \text{married}\}$.

h_t = health status at age t, $\{1 = \text{ good health}, 2 = \text{ bad health}, 3 = \text{dead}\}$.

hi_t = health insurance status (assumed to be time invariant), with $hi \in \{eph, gph, mca, nhi\}$ where these states are defined by:

eph: individual has employer-provided private health insurance, but no access to reasonably priced private health insurance (other than Medicare) if unemployed or employed part time

gph: individual has general private health insurance (such as Blue Cross/Blue Shield) that is not tied to employment at any particular firm,

mca: individual receives for Medicaid

nhi: individual does not qualify for Medicaid and does not have access to any type of fairly priced private health insurance.

aw_t = individual's Social Security Average Monthly Earnings (AME), discretized into *aws* categories. In this model there are $aws = 4$ equally divided intervals from the smallest to largest possible values of AME in the RHS sample.

Control variables:

ed_t = Employment decision. The version of the DP model estimated in this paper assumes that individuals have perfect control over their future employment status although they are uncertain about their realized future wage earnings from employment. This implies that $e_{t+1} = ed_t$ with probability 1.

ssd_t = Social Security application decision. This variable has two possible values, $\{1 = \text{apply}, 0 = \text{don't apply}\}$. This decision is not available for individuals who are under age 62 (since this model does not consider the DI application decision), or for individuals who are already eligible for Social Security OA benefits.

Before describing the estimation results, some comments on what we view are the most restrictive aspects of our specification of the state and control variables are in order.

1. The perfect control assumption implies that a) an unemployed individual who decides to return to work will be successful with probability 1 in obtaining a new job, and b) an employed individual who decides to quit their current job in the current period is pre-committed to sticking to that decision until the end of the following year. We simply note here that despite this short-term pre-commitment, the individual has 23 opportunities to reverse his decision. This is substantially more flexible than most retirement models which treat retirement as an absorbing state from which an individual is never permitted to exit.

2. Although one can also specify more realistic versions of the DP model that includes wealth and that allows for the consumption/savings decision such as described in Rust (1989), this version of the model treats consumption and income as identical: i.e. $y_t = c_t$ with probability 1. The rationale for doing this is twofold: a) consumption is very difficult to measure in the RHS dataset, and attempts to impute it from the budget constraint $c_t = y_t + w_t - w_{t+1}$ (where w_t is the individual's net worth at time t) yielded implausibly erratic consumption paths and a disturbingly high incidence of negative measured consumption, b) the predominantly blue-collar RHS sample has no significant tangible wealth beyond housing equity. Total net worth amounts to less than 5 years of income, and home equity consists of over 50% of this net worth. In view of the fact that most elderly appear averse to liquidating their housing equity to finance retirement consumption (Venti and Wise, 1990), the assumption that $c_t = y_t$ is not a bad approximation for the RHS sample. Indeed, the distribution of wealth changes in the RHS sample has a mean of \$$-658$ and a standard deviation of \$47,000, and a large peak at 0, so it would be hard to reject the hypothesis that $c_t = y_t$ if we assumed that any change in wealth $w_{t+1} - w_t$ is purely a result of measurement error. Indeed, a more detailed analysis of this issue in Rust (1990) reveals that most of the problems in imputing c_t from the budget equation result from errors in reporting wealth, especially from respondents who inconsistently reported certain components of their "balance sheet" which induced big errors in imputed consumption. In view of the potential measurement errors in wealth, we have opted not to include it as a state variable in the DP model, especially since we already include the individual's average monthly wage which is measured much more accurately and serves in many respects as a "permanent income" proxy for the individual's wealth. We also appeal to Cochrane (1989) who showed that the loss in utility of following the simple decision rule $c_t = y_t$ is negligible for an individual facing liquidity constraints, and Deaton (1991) who showed that setting $c_t = y_t$ is in fact an optimal strategy for individuals whose discount factors and income serial correlation coefficients are sufficiently close to 1.

3. Note that by defining income as net of health care expenditures we are implicitly assuming that individuals obtain no utility from health care *per se* but rather from their "stock" of health as represented by the h_t state variable, an interpretation that we don't view as being overly restrictive. However what is restrictive is the way the model deals with the event of having "negative income", i.e. when out of pocket health expenditures exceed current income. See section 5 for further discussion of this issue.

4. The final limitation, focusing on the retirement decision of male head of household rather than modeling the joint decision-making process of husband and spouse, is a simplification made primarily in the interest of computational tractability. Including the spouse as a decision-maker requires specification of additional state and control variables for the spouse's health, average wage, income, employment status, etc. The other rationale for avoiding the extra complexity of a joint retirement model is that dual careers were not an extremely common phenomenon for the RHS cohorts: spousal earnings are 0 in over 60% of male headed RHS households.

4. Estimation Results: Beliefs

Accurate modeling and estimation of individuals' beliefs about future mortality, health, marital status, wage earnings, health expenditures and Social Security benefits is the key to obtaining successful predictions from the DP model. Beliefs clearly involve many unobservable, subjective aspects, and consequently a number of strong assumptions must be invoked to estimate them. We also have to confront the problem of the huge number of possible future contingencies people can have beliefs about in the DP model. For example, even in our relatively coarsely discretized biennial version of the DP model, the x vector can take on 14,400 possible values at a given age t. Without further restrictions, the Markov transition matrix $p_t(x'|x,d)$ representing individuals' *one step ahead* beliefs is an array with 1.24×10^9 possible elements ($= 14400 \times 14400 \times 6$) and we must estimate 23 of these arrays for even values of age between 56 and 102. Obviously there is no possible way to estimate these arrays using only 7574 person-year observations without further *a priori* restrictions. The two main restrictions we employ are: 1) "rational expectations", i.e. individuals' subjective probability measures coincide with objectively estimable population probability measures, and 2) "exclusion restrictions". The latter are summarized in the following decomposition of p_t into a product of marginal and conditional densities:

$$
\begin{aligned}
p_t(x'|x,d) = & p_t^1(y'|y,ed,ss,ssd,m,h,aw)\times \\
& p_t^2(aw'|ed,y,aw)\times \\
& p_t^3(m'|m,h)\times \\
& p_t^4(h'|h,aw)\times \\
& p_t^5(\{h' = \text{dead}\}|h,m)
\end{aligned}
\tag{4.1}
$$

where p_t^1 is an income transition probability function, p_t^2 is a transition probability for the Social Security average wage, p_t^3 is a transition probability for marital status, p_t^4 is a transition probability for health status conditional on survival to the next period, and p_t^5 is the conditional probability of dying in the next period (mortality hazard function). Each of the probability functions $(p_t^1, p_t^2, p_t^3, p_t^4, p_t^5)$ can be estimated independently of each other in the first stage of the two-stage estimation procedure described in section 3. The marital status, health status, and mortality probabilities p_t^3, p_t^4, p_t^5 were specified as binary logit probabilities and estimated via maximum likelihood. The methods used to estimate the income and average wage transition probabilities p_t^1, p_t^2 are described in more detail in section 4.2 below. We do not have the space here to describe the estimation results in detail: interested readers should consult Rust (1996).

4.1 Health and Mortality

Individuals were classified as being in good or bad health based on a series of questions in the RHS survey, "do you have a health condition, physical handicap, or disability that limits how well you get around?" and "does your health limit the kind or amount of work or housework you can do?". If the respondent answered yes to one of these questions, we classified him as being in bad health, or in good health otherwise.[16] These sorts of self-reported health status measures have been criticized as endogenous on the grounds that people rationalize not being in the labor force by blaming it on poor health, leading to the presumption that the self-reported measures will overestimate of the impact of health on labor force participation (Myers, 1982, 1983 and Parsons 1982). However in a recent study comparing self-rated health to more "objective" measures of health status, Bound (1991) concluded that "Without further information it is not possible to determine whether objective or self-reported measures of health give more accurate indications of the importance of health in determining retirement behavior. . . . When outside information on the validity of self-reported measures of health are incorporated into the model, estimates suggest that the self-reported measures of health perform better than many have believed." (Bound, 1991, p. 17). We have experimented with other measures of health status in the DP model (including an alternative self-reported measure based on the response to the question "is your health better, worse, or the same as that of other people your age?") and have found that the results are fairly insensitive to the particular health measure used. However as we will see below, our simple binary health indicator is still a very imperfect measure of true health status.

Table 4.1 presents estimates of the estimated health transition probability $p_t^4(h'|h, aw)$ for different ages t and for the lowest and highest average wage classes aw. We exlude marital status as a conditioning variable since we found it to be statistically insignificant in predicting future health after conditioning on current health and the individual's average wage. The estimation results appear quite reasonable, with health status displaying a high degree of persistence, and the transition probabilities shifting in the expected direction with changes in age and average wages. We have also found that dummy variables for being on Social Security (particularly for being an early retiree) are significant predictors of future health status, even after conditioning on current health, age and average wage. The reason is clear from figure 2.2 of section 2: individuals who are in poor health are significantly more likely to apply for DI or early retirement benefits than those who are in good health. We view this finding as evidence that our simple binary health status indicator does not fully capture all dimensions of an individual's actual health status. Nevertheless

[16] A complication was caused by the fact that these questions were not asked in exactly the same form in the 1973 survey wave of the RHS, so some rather complicated adjustments were made to insure that health status was consistently defined.

we excluded Social Security status as a conditioning variable in p_t^4 on *a priori* grounds, since in the context of the DP model it has an obvious causal interpretation: applying for Social Security can be hazardous to your health.

Current Health Status	Current wage and Future Health Status			
	low wage		high wage	
	G	B	G	B
	Age 56			
G	80.2	19.8	91.5	8.5
B	34.3	65.7	58.1	41.9
	Age 64			
G	71.6	28.4	87.0	13.0
B	24.5	75.5	46.4	53.6
	Age 80			
G	49.4	50.6	72.2	27.8
B	11.2	88.8	25.1	74.9

Table 4.1 Examples of Health Status Transition Probabilities

Figure 4.1 presents estimates of the mortality hazard function (on an annual basis), which varies substantially depending on the person's age, health and marital status. Individuals who are single or in bad health are much more likely to die than individuals who are married or in good health. A check on the accuracy of the mortality model is to compare the mortality rates for the four subgroups (married/good health, married/bad health, single/good health, and single/bad health) with mortality rates for males from 1970 U.S. Census projections. The Census mortality rates are not conditioned on health and marital status, so we computed the average RHS mortality rate by taking the weighted average mortality rates for the four subgroups and compared it to the Census mortality rate for all males of a given age. As you can see, the estimates from the DP model agree quite closely with the Census estimates. A limitation of the RHS is that it contains no data on the very old: the oldest male in the sample is 73. Since the DP model requires estimates of mortality rates to age 102, we extrapolated our estimates from ages 74 to 102 in such a way that the weighted average of the four curves matched the Census projections, where the weights are based on our projections of the sample proportions of individuals in the four health/marital states (as predicted from the Markov chain models for health and marital status discussed above). Thus, while the close correspondence between the mean hazard rate and the Census projections over ages 56 to 73 can be viewed as a validation of the accuracy of our estimates, the correspondence between the two curves between ages 74 to 102 is entirely by construction.

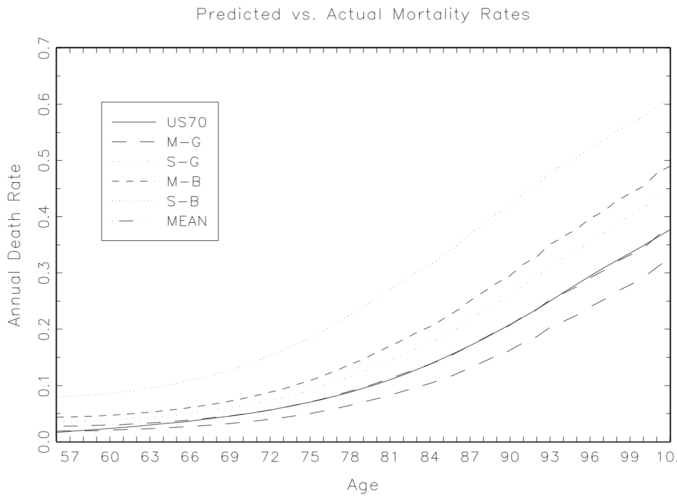

Figure 4.1 Mortality Rates for RHS Males by Health and Marital Status

4.2 Wage and Asset Income

The income transition probability matrix p_t^1 is the key component of the DP model since income is the main determinant of individuals' future utility, and accurate modeling of how future income depends on current decisions is critical to the DP model's ability to mimic observed retirement behavior. Even after making the exclusion restrictions in equation (4.1), the p_t^1 component amounts to some 248,400 different income distributions (equal to the total number of possible conditioning cells in p_t^1, $t = 56, \ldots, 102$ over the remaining 23 periods of life). So once again we adopt the strategy of decomposing p_t^1 into product of individual sub-transition matrices. Our decomposition specifies the distribution of total income y_t as a convolution of separate distributions for the following components,

$$y_t = \mathit{rw}_t \otimes \mathit{sw}_t \otimes \mathit{rss}_t \otimes \mathit{sss}_t \otimes \mathit{oi}_t \otimes \mathit{hc}_t \tag{4.2}$$

which amounts to an assumption that the individual components of total income are conditionally independent of each other. Actually, wage income and Social Security income are not conditionally independent in our specification due to the operation of the earning test provision of the Social Security law. The dependence is captured through the use of "Social Security sub-transition matrices" $\pi(\alpha)$ that will be described in more detail in section 4.3.

Excluding Social Security, estimation of p_t^1 under either of the decompositions outlined above requires estimation of three further sets of income distributions: wage earnings of the male, wage earnings of the spouse, and asset earnings. The most important component is the male's distribution of wage earnings since spouse wage earnings and asset income constitute a relatively insignificant source of total family income for this subsample. For example, for the RHS as a whole, spouse earnings amounted to only 14% of total family income on average, and constituted 0% of total income for more than 60% of all households. The distribution of asset and other non-wage, non-pension income was even more concentrated at 0. Therefore we used a two-stage procedure to estimate the conditional distributions of spouse wage and asset income: 1) we estimated separate binary logit models to predict the probability that spouse and asset income are 0 conditional on a set of observed state variables, and 2) we computed non-parametric (histogram and kernel) density estimates of spouse and asset income conditional on the other state variables and the event that next period income is greater than 0. We also estimated parametric lognormal models and found that they provided relatively good approximations to our non-parametric estimates of the income distributions. Our estimate of the overall distributions of spouse wage and asset income is then simply a mixture of the mass point at 0 (predicted by the logit model) and the estimated conditional income densities over the positive real line.

We devoted considerable care to estimating the distribution of male wage earnings, since wage income together with Social Security benefits constitute the bulk of total family income and is therefore a key driving force behind the DP model. We estimated conditional distributions of the form $f_t(rw'|ed, y, aw, h, m, hi)$ where rw' is the male's predicted labor income at time $t+1$. Initially we estimated f_t by a non-parametric histogram approach, with local averaging across conditioning cells that contained fewer than 50 observations. Similar to spouse and asset income, we found that the non-parametric estimates of the wage distributions were well approximated by log-normal distributions. We subsequently discovered an even better approximation: the *ratio* of actual wage earnings y_{t+1} to the individual's average monthly wage aw_t has a distribution that is very close to log-normal. Table 4.2 presents estimation results of a regression of the ratio y_{t+1}/aw_t on various independent variables including the individual's age t, health status, h_t, marital status m_t, Social Security status, ss_t, health insurance, hi_t, and a series of dummy variables interacting hours of work ed_t with the individual's average wage aw_t (the latter interactions are not reported in Table 4.2). The estimated parameters seem quite reasonable: the distribution of wages drifts downward with age, is affected negatively by being in poor health, and is significantly lower for early retirees and individuals with no health insurance. We do acknowledge that the negative coefficients for the Social Security dummies may be partly spurious due to a failure to control for unobserved heterogeneity since our controls for "observed heterogeneity" such as h_t and aw_t are imperfect measures of the individuals' true health and earning potential. In particular, early retirees tend to be lower wage earners than normal retirees and also tend to have more physically demanding jobs. Thus, problems of selectivity bias may arise if unobservables connected with the decision to retire or continue working are correlated with unobservable factors affecting labor income. However we think a more probable explanation for the significance of these dummies is that it reflects a problem with our relatively coarse discretization of hours of work: a significant fraction of early retirees continue working on a full or part-time basis to supplement their Social Security benefits, reducing their hours of work enough so that their Social Security benefits are not taxed away by the Social Security earnings test. Rather than estimate a more detailed DP model with a finer discretization of hours of work, we chose to retain our 3 state (FT/PT/NE) discretization and incorporate the disincentive effects of the Social Security earnings test indirectly via these "endogenous income distributions" which we interpret as reflecting workers' optimal responses to the Social Security earning test. Although it would be preferable to make this shift in earnings a result of an explicit choice of reduced hours of work within the DP model, we do not believe it leads to spurious conclusions about the disincentive effects of the earning test tax.[17]

Figure 4.2 compares the non-parametric estimate of the estimated residuals to the normal distribution, which indicates that the log-normal specification fits the data reasonably well. Using the estimated log-normal model, it is then straightforward to compute p_t^1 by integrating the probability that wage income falls in each cell, providing a flexible way of constructing discretized income distributions with an arbitrary number of cells.

Parameter	Estimate	Standard Error	t-statistic
$h_t = B$	-.044	.013	-3.3
$m_t = M$	.012	.23	0.5
$t = 59$	-.034	.032	-1.0
$t = 60$	-.062	.028	-2.2
$t = 61$	-.043	.028	-1.5
$t = 62$	-.043	.027	-1.6
$t = 63$	-.048	.028	-1.7
$t = 64$	-.070	.030	-2.3
$t = 65$	-.138	.036	-3.9
$t \geq 66$	-.249	.037	-6.7
$(t-66)/(1+t-66)$	-.330	.036	-9.1
$hi_t = eph$	.036	.028	1.3
$hi_t = gph$	.056	.034	1.6
$hi_t = nhi$	-.105	.035	-3.0
$ss_t = 62$	-.358	.050	-7.1
$ss_t = 63$	-.296	.074	-4.0
$ss_t = 64$	-.198	.057	-3.5
$ss_t = 65$	0.04	.054	0.7
$m_t = M, ss_t = 62$	.051	.052	1.0
$m_t = M, ss_t = 63$	.064	.079	0.8
$m_t = M, ss_t = 64$	.070	.058	1.2
$m_t = M, ss_t = 65$	-.039	.053	-0.7
$\hat{\sigma} = 0.704$	$R^2 = .3589$	$N = 10040$	

Table 4.2 Regression estimates for male log-wage ratio $log(y_{t+1}/aw_t)$

[17] A previous version of this model also estimated a specification with “exogenous income distributions”, i.e. where the Social Security dummies were excluded. The predictions and fit of this model are basically the same as the version with endogenous income distributions.

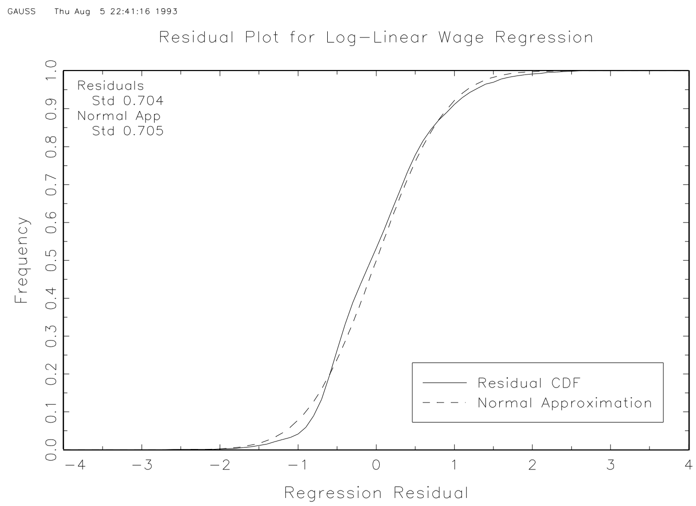

Figure 4.2 Empirical vs. theoretical CDF's of residuals from regression on male log-wage ratios

4.3 Social Security Benefits

After the male's wage earnings, Social Security benefits are the next most important source of family income in our sample. This section describes how we embedded the Social Security rules into the income transition probability p_t^1. Before describing the rules it is important to note that our model treats Social Security policy α as time-invariant whereas in fact several changes in Social Security rules and benefit levels were enacted over the 1969-1979 period of our sample. Treatment of changes in Social Security policy introduces substantial additional complexities, especially if we want to model policy as a stochastic process $\{\alpha_t\}$. This is a task we leave to future research. For our purposes here, it is important to observe that although there were important changes in Social Security benefits in 1967, 1969 and 1971, policy changes "stabilized" with the automatic indexation of benefits in the 1972 Social Security amendments.[18] It is convenient to assume, therefore, that individuals in our sample (who were between 58 and 63 at the beginning of the RHS survey in 1969) correctly anticipated the 1972 Social Security amendments and the large real benefit increases that occurred during the early seventies, and also that they also correctly anticipated that they would be grandfathered under the 1977 amendments and therefore protected from the large relative decreases in benefits targetted for subsequent cohorts.

The rational-expectations/perfect foresight assumption is a testable hypothesis: the 1969, 1971 and 1973 waves of the RHS asked respondents what they expected their Social Security benefit levels would be at the time they retired. Bernheim (1988a,b) found that although reported expectations were rather noisy estimates of actual future benefit

[18] The 1972 amendments actually contained an error that lead to an inadvertent double indexation of benefits that was rescinded in the 1977 Social Security amendments. However people born prior to 1917 (which includes all the RHS sample) were grandfathered under the 1977 amendments. This grandfathering clause is fortunate for our purposes since there was a significant decline in real Social Security benefits (relative to what benefits would have been had the amendements not been enacted) for the "notch baby" cohort born between 1917-1921. See Kreuger and Pischke (1991) for an analysis of the effect of the 1977 legislation on the labor supply decisions of the notch babies.

levels, "When one corrects for the presence of reporting error through the appropriate use of instrumental variables, the resulting estimates are generally consistent with the theory (of rational expectations). In particular, one cannot reject the hypotheses that expectations evolve as a random walk, and that innovations in this process are unrelated to prior information." (p. 4). Furthermore, he found that the quality of individuals' reported expectations improved as they approached retirement age: "The results are striking. Responses to new information during the period immediately preceding retirement appear to be highly rational" (p. 4). Of particular interest for our purposes is Bernheim's conclusion that "the 1972 legislation was largely anticipated, and the summary statistics in Table 1 show little evidence of an upward surge in expectations after 1972." (p. 18).

These results provide some justification for our assumption that individuals' beliefs about future Social Security benefits reflected a correct forecast of the Social Security law as it existed following the 1972 amendments. The main features of these rules can be briefly summarized as follows:

1. PIA is a nonlinear function of AME as shown in figure 4.3 where the wife's benefit is the larger of 50% of the male's PIA or her own PIA computed from her own earnings history, and the overall monthly benefit is no higher than the maximum family benefit level specified in the *Social Security Handbook,*
2. AME is the average of *nominal* wages earned during individual's *computation years* divided by the number of months in those years. The computation years are the N years of highest earnings from 1950 until the year the individual is entitled for benefits and N is the number of years from 1950 until the individual turns 57.
3. the (permanent) benefit reduction factor for retirement prior to age 65 is 5/9 of 1% for each month of entitlement to benefits prior to age 65;
4. a DRC (delayed retirement credit) of 1/12 of 1% is paid for each month retirement is delayed beyond age 65;
5. Social Security benefits are subject to an earnings tax of 50% of earnings in excess of the test level approximately equal to $2,000 in 1968 dollars (except that the tax rate is 0% for individuals over age 72),
6. Medicare insurance is available to individuals who are over 65 and are entitled to Social Security old age benefits.

We now show how one can approximate these rules via a vector of policy parameters α and embed the rules as component of the (controlled) Markov transition probability for income p_t^1 and the aw transition matrix p_t^2. The actual rules are clearly not first order Markovian due to the fact that aw_t is a moving average of the entire history of earnings, and not just the most recent value. However table 4.3 shows that not withstanding the double-indexation that occurred in the 70's, by the time one reaches retirement age there is very little variation in aw_t. Therefore we approximate its transition matrix, p_t^2, by an identity matrix. This implies that aw is a time-invariant person-specific effect.[19]

[19] The results presented in table 4.3 are an average over both workers and non-workers. If we condition on just the sample of workers, we see more significant changes in aw_t from period to period. In future versions of our DP model we plan to incorporate the effects of AME "bracket creep" by embedding separate AME transition matrices into p_t^1 depending on the individual's income and hours of work.

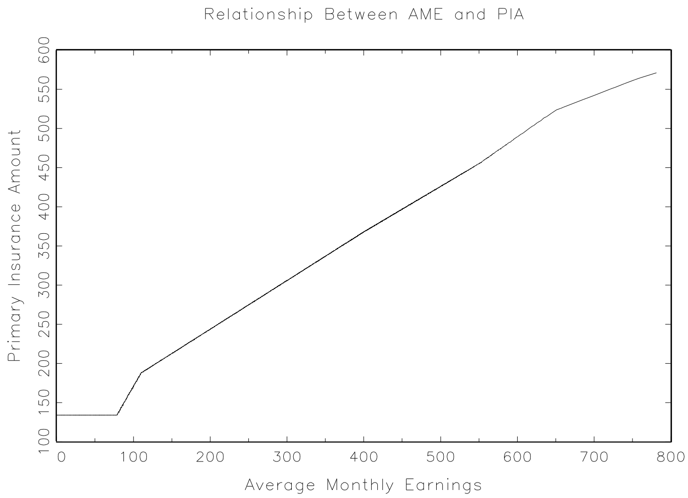

Figure 4.3 Relation between AME and PIA under the 1972 Social Security Amendments

$t/t+1$	$aw'=1$	$aw'=2$	$aw'=3$	$aw'=4$	$aw'=5$	$aw'=6$	$aw'=7$	$aw'=8$
$aw=1$	98.34	1.66	0.00	0.00	0.00	0.00	0.00	0.00
$aw=2$	0.00	97.95	2.05	0.00	0.00	0.00	0.00	0.00
$aw=3$	0.00	0.00	97.77	2.23	0.00	0.00	0.00	0.00
$aw=4$	0.00	0.00	0.00	97.26	2.74	0.00	0.00	0.00
$aw=5$	0.00	0.00	0.00	0.00	96.82	3.18	0.00	0.00
$aw=6$	0.00	0.00	0.00	0.00	0.00	97.28	2.72	0.00
$aw=7$	0.00	0.00	0.00	0.00	0.00	0.00	97.65	2.35
$aw=8$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Table 4.3 AME Transition matrix for 1974

The relation between AME and PIA and the earnings test tax (parts 2 and 3 of the Social Security rules listed above) are encoded via a set of $ys \times ys$ *replacement rate* matrices $\pi(y'|y, \mathit{aw}, t, m, \mathit{ss}, \alpha)$, which are also conditioned on the individual's Social Security average wage, age, marital status, and Social Security state (i.e. the age of initial eligibility for Social Security). The π matrix transforms the discretized distribution of wage income of the husband and spouse prior to Social Security benefits into a distribution of wage income including Social Security benefits. The form of these transition matrices fully embody the Social Security rules for determination of OA benefits, and in particular reflect the impact of the Social Security earnings test as illustrated in figure 4.4. The left panel of figure 4.4 presents the π matrix for a single individual of age 66 who first became eligible for Social Security benefits at age 62 with an average monthly earnings of $265.81 (corresponding to a PIA of $284.60). The transition matrix shows that if the individual does not work at all, he can expect a Social Security benefit of approximately $2,700 (equal to 12 times the individual's PIA times the early retirement benefit reduction factor of .8). As the individual's wage income increases, his combined wage and Social Security income line in figure 4.4 parallels the diagonal line (representing wage income alone) until the worker's wage income exceeds the earnings test level of $2,000. From that point on, Social Security benefits are reduced by 50 cents for every additional dollar of earnings until at an earnings level of approximately $7,200, 100% of the individual's Social Security benefits have been taxed away and the post-Social Security benefit line coincides thereafter with the diagonal no benefits line in the π matrix. The right hand panel of figure 4.4 presents the π matrix for a married 72 year old individual with a higher average monthly wage. This individual expects a combined husband and spouse Social Security benefit of $5,000 per year. Notice that unlike the left panel, the post-benefit line is parallel to the diagonal pre-benefit line. This is a result of the Social Security rule that stipulates that persons 72 or older are not not subject to the earnings test tax on post-retirement wage earnings. In general, variations in Social Security rules such as the level of the reduction factor for early retirement benefits, the shape of the schedule relating AME to PIA, etc. are completely embodied in the π matrices. The only other component of the p_t matrices that changes as a result of changes in Social Security policy is the transition probability matrix for net out-of-pocket health care costs, which reflects the impact of Medicare insurance reimbursement policies such as the age of eligibility for Medicare, the co-payment level, etc. All remaining components of beliefs such as the distribution of husband and spouse wage earnings, the transition probabilities for health and marital status, and the mortality hazard, are assumed to be invariant to changes in Social Security policy.

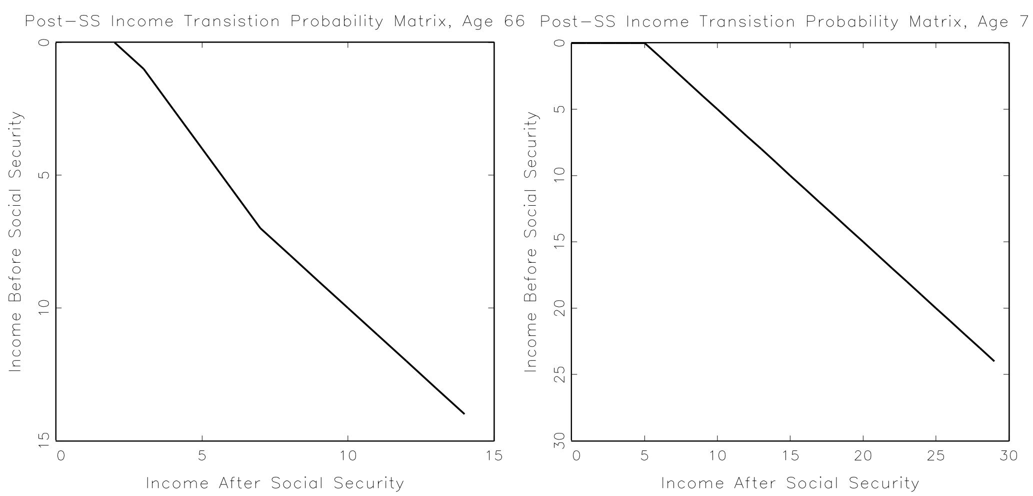

Figure 4.4 Embedding Social Security Earnings Test Rules as a Transition Matrix: Ages 66 and 71

4.4 Health Expenditures

As we discussed in the introduction, it is important to incorporate people's uncertainty over health care costs and insurance coverage to enable the DP model's to capture differential retirement patterns for individuals with different access to private health care insurance, an effect that we conjecture is partly responsible for the large observed peak in retirements at age 65. The DP model "rationalizes" these findings as result of the interaction of risk-aversion, incomplete private health insurance markets, and the "Medicare insurance option". However to model people's subjective beliefs about health care risks, we either need a substantial battery of questions to elicit what those beliefs are, or we must impose the hypothesis of rational expectations and estimate people's beliefs from data on their actual health care expenses, both pre and post insurance reimbursement. The RHS collected such information, in the form of questions asking respondents how much their total medical bills were in the previous year, how much much of that total was reimbursed by insurance, and how much was paid out of pocket. The DP model also requires us to determine what types of health insurance each individual has access to. Appendix 1 describes how the available information from the RHS was used to infer the types of health insurance an individual has access to based on their employment status and their responses to questions about private health insurance and Medicaid coverage. This information allowed us to classify each respondent into one of four mutually exclusive health insurance opportunity sets *eph*, *gph*, *mca,* and *nhi* defined in our summary of state variables in section 3.

The information on health insurance and expenditure and the hypothesis of rational expectations allow us to specify a model of individuals' beliefs of the effects of having various combinations of private and public health insurance, both before and after insurance reimbursement. Plotting the distribution of health care expenditures, the two most striking features are the high probability of small health care expenditures (i.e. less than \$500), and the extremely long tail representing the "rare event" of a large, catastrophic health care expenses. Similar to the estimation procedure for spouse and asset income, we treated the distribution of health care expenses as a mixture of a mass point on expenditures of \$500 or less, and a continuous distribution of the conditional distribution of health care expenses in excess of \$500. Interestingly, we discovered that the upper tail of the health care distribution is almost perfectly described by the Pareto distribution

$$p_t^5(y|\mathit{hi}, h, m) = \frac{\gamma 500^\gamma}{y^{\gamma+1}} I\{y \geq 500\}. \tag{4.3}$$

This is verified in figure 4.5 which compares the estimated Pareto distribution, and two non-parametric density estimates: a simple histogram estimate and a kernel density estimate using a lognormal kernel. The very low value of the Chi-square goodness of fit statistic relative to its degrees of freedom ($\chi^2 = 559, df = 689$) shows that there is little evidence against the simple Pareto specification as is readily apparent from figure 4.5. The Pareto distribution provides a convenient way to assess the adequacy of various forms of health insurance, since it is characterized by a single parameter γ and higher values of γ correspond to distributions with smaller tails (i.e. lower health care risks).

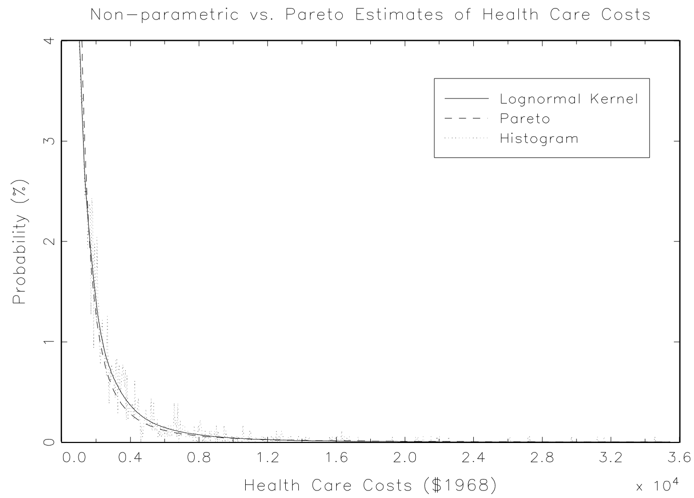

Figure 4.5 Non-parametric vs. Pareto Density Estimates of the distribution of health care costs

The maximum likelihood estimates for total health care costs (excluding health insurance premia) is a Pareto with $\gamma = 1.04$, which implies a rather large potential risk of catastrophic health care costs since the conditional mean of the upper tail of the Pareto distribution (i.e. conditional on health care expenditures exceeding \$500) is $\gamma 500/(\gamma - 1)$ which equals \$13,000 in this case. The estimates show that health insurance does not function perfectly in the sense of removing all health care risks and replacing them by a fixed premium equal to the expected value of health care costs. Instead insurance partially reduces health care risks by increasing the value of γ.

Overall, the estimates of the effects of different types of insurance coverage are quite reasonable, corresponding to what we know *a priori* about the relative generosity of various forms of health insurance coverage. The maximum likelihood estimate of γ for individuals who only have Medicare insurance is $\hat{\gamma} = 1.76$, for those with private health insurance $\hat{\gamma} = 2.29$, and for those on Medicaid $\hat{\gamma} = 2.33$. The estimated γ for individuals who have both Medicare and private health insurance is $\hat{\gamma} = 2.37$, reflecting in part the complimentary nature of "Medigap" insurance policies. [20] Note that our estimates of the distribution of health care expenses net of insurance reimbursements implicitly account for health insurance premiums via the estimated logit probabilities that health care costs under \$500. Thus, the estimated distributions reflect the fact that the higher quality coverage provided by private health insurance plans comes at the price of higher premiums, which are reflected in a substantially lower estimated probability of having health care expenditures less than \$500. On the other hand Medicare and Medicaid have much poorer coverage of large health care expenses than private health insurance, but have relatively minimal premiums, deductibles, and co-payments for covered services which are reflected in a substantially higher probability of having health care expenditures of \$500

[20] It is interesting to note that the estimated γ for Medicaid is significantly higher than for Medicare. Medicaid provides broader coverage of health care problems than Medicare, although government mandated maximum allowable hospital and physician charges are lower under Medicaid than Medicare.

or less. Our estimates indicate that private health insurance provides better coverage of catastrophic health risks than does Medicare (as evidenced by the higher estimated γ for the former), but at the cost of substantially higher annual premiums.

5. Estimation Results: Preferences

Having described the DP model and our procedure for estimating individuals' beliefs, we now present a simple parametric specification of individuals' preferences and the maximum likelihood estimates of the unknown parameters. Of the many possible specifications of preferences, we have focused on the additive-separable family

$$u_t(y, e, \mathit{ed}, h, m, \theta_u) = u_t^1(y, h, m, \theta_u^1) + \sum_{i=1}^{\mathit{es}} \sum_{j=1}^{\mathit{es}} u_t^2(i, j, h, m, t, \theta_u^2) I\{e = i, \mathit{ed} = j\} \tag{5.1}$$

According to this specification, the function u_t^1 represents the utility of consumption (which equals income in this model), which is shifted by age, marital and health status. The function u_t^2 represents the utility of various possible labor/leisure/search decisions which can also depend on age, health and marital status. The double summation in (5.1) reflects the various possible transitions an individual can make, $\mathit{es} = 3$ in the current specification. Recall that $e = \mathit{NE}$ denotes unemployment/out of the labor force, $e = PT$ denotes part-time employment, and $e = FT$ denotes full-time employment. Using a more suggestive notation and excluding the other arguments of u_t^2 for clarity, $u_t^2(\mathit{NE} \rightarrow \mathit{NE})$ represents the utility of leisure of an unemployed individual who chooses to remain unemployed, $u_t^2(\mathit{NE} \rightarrow FT)$ represents the potential disutility of search costs incurred by an unemployed individual who chooses to return to work full-time, $u_t^2(FT \rightarrow \mathit{NE})$ represents the utility of a full-time worker who chooses to quit his job, and so on. *A priori* we expect individual to prefer leisure to work, so that for any current labor force state e we have

$$u_t^2(e \rightarrow \mathit{NE}) > u_t^2(e \rightarrow PT) > u_t^2(e \rightarrow FT) \tag{5.2}$$

Our specification allows the relative values of leisure and work to depend on the current labor force state, e. For example, the disutility of working full-time relative to not working may be much higher if the individual is currently unemployed and must incur monetary and psychic costs of searching for a new job than if the individual is currently employed. The specification in (5.1) also allows these relative utilities to depend on the individual's current age, health, and marital status. For example, we would expect that the disutility of not working relative to working would be much higher for an unhealthy individual than a healthy individual. It is less clear how the utility of work should depend on the individual's age and marital status, although it is certainly reasonable to expect that the utility of leisure should increase with age, and should be higher if the person has a spouse to share their leisure time with than if they were single. However in the versions of the DP model we have estimated to date, we have excluded health, age and marital status as arguments to u_t^2, estimating this function simply as the coefficients on the 9 dummy variables, $I\{i = e_{tk}, j = ed_{tk}\}$, $i, j \in \{FT, PT, NE\}$, where e_{ti} and ed_{ti} are the observed states and decisions of individual

k at age t. A major reason why we have adopted this more Spartan specification of utility is to determine whether Social Security incentives alone are sufficient to explain the observed pattern of retirement behavior. We know that we can capture the large peaks in retirements at ages 62 and 65 by including age-specific dummies. Similarly we can capture the higher likelihood of early retirement of individuals who are in poor health by including dummies that makes the disutility of working very high. However resorting to age dummies is clearly an unsatisfactory way to "explain" the peaks in retirements at 62 and 65: it begs the question of why people should happen to have especially strong preferences for leisure at precisely these ages.

Our *a priori* expectations for the utility of income function u_t^1 are that it should be increasing and concave in income y reflecting risk aversion on the part of the individual. However we have no strong *a priori* beliefs about how u_t^1 should depend on age t, health h_t, and marital status m_t. Rather than completely exclude these variables from the utility function, we estimated models where u_t^1 is a member of the constant relative risk aversion family:

$$u_t^1(y,h,m,\theta_u) = \left[\frac{y^{\theta_{11}}}{\theta_{11}}\right] exp\left\{\theta_{12} + \theta_{13}I\{h = \textit{bad}\} + \theta_{14}I\{m = \textit{married}\} + \theta_{15}t/(1+t)\right\}, \tag{5.3}$$

where the utility of income is shifted multiplicatively by an exponential factor containing dummy variables for being in bad health, being married, and age. We imposed an *a priori* restriction that age should have a smooth, monotonic effect on the utility of income via the logistic specification $t/(1+t)$, to ensure that the DP model's ability to fit the data is not a result of inclusion of *ad hoc* age dummies.

However after some experimentation, we found that we could substantially improve the DP model's ability to fit the data by incorporating two admittedly *ad hoc* features: 1) allowing for a substantial disutility of receiving "negative income", and 2) including dummy variables that allow the disutility of working to depend on the individual's Social Security status. We explain the rationale for each of these features in turn:

1. Recall that in our model consumption is assumed to equal income net of health care expenditures in each period. We excluded wealth as a state variable and the consumption/savings decision as a control variable due to large measurement error problems that we discussed in section 4. Since we haven't included wealth and haven't explicitly modeled the event of personal bankruptcy, we need some way of accounting for what happens in the event that health care costs (after insurance reimbursements) exceed total income in a given year. We model this event by the artifice of "negative income", and use dummy variables to let the data tell us exactly how averse individuals are to such an event. To exclude these dummy variables would amount to an implicit assumption that there is an unamed, costless source of "health insurance of last resort" that will always be able to cover any shortfall between income and net health care expenses. While it is certainly true that some individuals – especially indigent individuals without health insurance – seem to have some sort of insurance of last resort (typically due to hospitals' decisions to treat these patients and use "cost-shifting" to effectively

force Medicare, Medicaid, and private health insurance to cover their bill), there is good reason to believe that most individuals are highly averse to being in a situation where they are unable to pay their bills (due to stigma, fears of receiving delayed or inferior health care, etc.). Rather than excessively complicating the DP model by attempting to explicitly model the somewhat murky issues of what happens when health care bills exceed one's current income, we simply included a set of dummy variables for the event net income is negative in order to let the data speak to the issue of exactly how averse individuals are to such an event. The negative income dummies were interacted with marital status and the individual's average wage on the grounds that average wage is a proxy for permanent income (wealth), and individuals with higher wealth should be more concerned with having inadequate insurance coverage since they have more wealth at stake to lose.

2. We included dummy variables for individuals who are receiving Social Security benefits and continuing to work part-time or full-time as a way of testing whether our three way discretization of hours of work discussed in section 5.3 was overly coarse. The significance of these variables suggests that the discretization is indeed too coarse and that finer discretizations should be used in the future. The idea is that individuals who are receiving Social Security may reduce their hours or work effort so their wage earnings complement their Social Security benefits but do not substantially exceed the $2,000 earnings test level where they begin to face a 50% wage surtax. Given our coarse 3-way discretization of hours of work, there is no way for this reduced work effort to be directly reflected in the model under our current specification except through these dummy variables which can be interpreted as leisure "bonuses", i.e. reductions in the disutility of full or part-time work as a result of reduced total hours of work in each of these categories. Clearly a better way to handle the problem is to allow for a finer discretization of hours of work, an approach we will pursue in future work.[21] In addition to Social Security, we included interaction terms of the form $I\{ed = FT, e = FT, aw = j\}$, $j = 1, \ldots, aws$ to capture the possibility that high wage professional workers may have better working conditions leading to lower disutility of continued work than lower wage blue collar workers.

With these caveats in mind, we are now ready to turn to the parameter estimates in table 5.1. The table presents estimates of of the utility function with discount factor fixed at $\beta = .98$. We find that the estimation results are quite reasonable, corresponding to our *a priori* expectations. In particular, the coefficient of relative risk aversion, θ_{11}, is relatively precisely estimated at $-.07$, reflecting significant risk aversion on part of these individuals as expected. The estimates of the u^2 function also conform to our prediction in (5.2) that in each state e, the utility of not working is greater than the utility of working part time, which is in turn greater than the utility of working full time. We also

[21] Excluding the Social Security dummies from the utility function does not result in a major deterioration in the ability of the model to fit the data. In particular, these variables do not represent a sly attempt to enable the model to capture peaks at ages 62 and 65. Note that these dummies are not interacted with age and apply only to the relatively small subset of individuals who decide to continue working after applying for Social Security benefits. None of the results in the subsequent sections would change qualitatively if these variables were omitted from the specification.

find that the disutility of full-time work relative to leisure for currently unemployed males $(-12 = (-65) - (-53))$ exceeds the disutility of full-time work relative to leisure for currently employed males $(-7 = (-54) - (-47))$, as expected. Although the differences in these coefficients do not appear to be significant relative to their estimated standard errors, since these are uncorrected standard errors we are unwilling to engage in formal hypothesis testing at this point.[22] Although the standard errors may be large, we can say unambiguously that these parameters are very important for enabling the DP model to fit the data: even relatively small changes in individual coefficients yield sharp degradations in the likelihood function. We have already noted that there is an implicit multicollinearity problem that contributes to the large standard errors: the 9 dummy coefficients comprising the u^2 component of the utility add up to a constant function which is evidently highly collinear with the sum of the u^1 component and next period value function v_{t+1} (which provides the expected discounted value of utility of income). If we impose an *a priori* normalization of one of these coefficients to 0 (re-computing the maximum likelihood estimates of the remaining 8 dummy variables) the standard errors on the u^2 coefficients become much smaller, although the likelihood drops significantly.

We also find that the large negative coefficients for the negative income dummies reflect substantial aversion to the event that net health care expenditures exceeds total income, especially for individuals with higher average wages. This most likely reflects the fact that the higher wage individuals have more net worth and thus more at stake to lose in the event that health care expenses exceed income, providing a potential explanation for the observation that the likelihood of health insurance coverage is an increasing function of wages and net worth. Finally, the positive coefficients on the ed, ss interaction terms confirms our expectation that these coefficients would represent utility "bonuses" reflecting a lower disutility of full or part time work by individuals who are receiving Social Security benefits, a finding consistent with the observation that some individuals who continue to work on a full-time or part-time basis while receiving Social Security benefits have reduced their hours work so that they are nearer the $2,000 earnings test threshold where the 50% surtax kicks in. The only "surprising" results are the findings that individuals in bad health have higher marginal utilities of consumption than those in good health and that neither marital status nor age have a significant effect on the utility of consumption.

Due to the fact that the likelihood is very flat as a function of β, we did not attempt to include it as part of the vector of utility function parameters θ in estimation. Instead we estimated the subjective discount factor β via grid search. The precise estimate of β depends on the particular utility specification considered, but in all the models we examined, the maximum likelihood estimate of β was very close to 1, indicating that individuals are very far-sighted with low subjective discount rates. The results in table 5.1 correspond to the value of β that maximized the likelihood function, which corresponds to a (continuously compounded) discount rate of 1.01%.

[22] Recall from section 3 that if the full information matrix is block diagonal, then the estimates provided in table 5.1 are fully efficient and the standard errors are consistently estimated.

Parameter	Estimate	Standard Error	t-statistic
θ_{11} (risk aversion)	-.072	0.034	-2.1
θ_{12} (constant)	2.696	0.899	3.0
θ_{13} (bad health)	0.235	0.089	2.6
θ_{14} (married)	0.040	0.102	0.4
θ_{15} (age logit)	-1.637	1.04	-1.6
$u^2(FT \to FT)$	-54.72	20.08	-2.7
$u^2(FT \to PT)$	-51.80	20.22	-2.6
$u^2(FT \to NE)$	-47.24	20.84	-2.3
$u^2(PT \to FT)$	-60.77	20.04	-3.0
$u^2(PT \to PT)$	-56.00	20.11	-2.8
$u^2(PT \to NE)$	-51.99	20.74	-2.5
$u^2(NE \to FT)$	-65.44	19.32	-3.4
$u^2(NE \to PT)$	-60.61	19.42	-3.1
$u^2(NE \to NE)$	-53.39	19.99	-2.7
$I\{y \leq 0, m = M\}$	-9.72	6.68	-1.5
$I\{y \leq 0, aw = 1\}$	0.236	7.35	0.0
$I\{y \leq 0, aw = 2\}$	-5.64	7.93	-0.7
$I\{y \leq 0, aw = 3\}$	-30.29	7.68	-3.94
$I\{y \leq 0, aw = 4\}$	-92.44	33.94	-2.7
$I\{e = FT, ed = FT, aw = 2\}$	0.124	0.077	1.6
$I\{e = FT, ed = FT, aw = 3\}$	-0.082	0.083	-1.0
$I\{e = FT, ed = FT, aw = 4\}$	0.501	0.162	3.1
$I\{ed = FT, ss = ER\}$	0.372	0.054	6.8
$I\{ed = PT, ss = ER\}$	1.767	0.099	17.7
$I\{ed = FT, ss = NR\}$	0.876	0.055	15.8
$I\{ed = PT, ss = NR\}$	1.907	0.103	18.5
$L(\hat{\theta}) = -5953.94$	$\beta = .98$	$N = 7574$	

Table 5.1 Utility Function Estimates

6. Comparison of Predicted vs. Actual Retirement Behavior

One can evaluate the DP model along several dimensions, on intuitive grounds such as the economic plausibility of the estimated preferences and beliefs, and on statistical grounds such as its ability to pass formal goodness of fit tests. An even more rigorous evaluation would be to test the DP model's ability to predict behavior out-of-sample, ideally under under new environments and Social Security policies, a task we are currently working on. However there is no point in worrying about out-of-sample forecasting if the model is implausible, or is unable to even fit the data in-sample. The economic plausibility of the estimated preferences and belief parameters has been discussed in previous sections. The remainder of this section provides some evidence on the ability of the DP model to account for observed behavior, and to provide insight into the behavioral impacts of Social Security and Medicare.

For reasons discussed below, we do not conduct formal statistical tests of goodness of fit. Instead we directly compare the parametric estimates of the conditional choice probability function $P(d|x,\hat{\theta},\alpha)$ to non-parametric estimate $\hat{P}(d|x)$ which can be thought of as summarizing individuals' actual behavior patterns (decision rules). Given the discreteness of the state and control variables, the non-parametric estimate $\hat{P}$ is simply the sample histogram of choices made by the subsample of individuals whose state is x. Note that DP model has $345,600$ possible x cells over the age range covered by the RHS, whereas our data set of 7574 person/year observations only has information on 2880 distinct x cells, with an average of only 2.63 observations per cell. The NP model must estimate 5 unknown parameters in most of these cells (equal to the number of possible choices less 1), so it is clear that comparisons of cell-specific NP estimates to the predictions of the DP model are not meaningful unless aggregated over larger collections of x-cells. Given a collection A of x cells, it is possible to compute the parametric and non-parametric estimates of $P(d|A)$ by sample enumeration:

$$\begin{aligned}\hat{P}(d|A) &= \int_{x\in A}\hat{P}(d|x)\hat{F}(dx|A) \quad \text{(NP)}\\ P(d|x,\hat{\theta},\alpha) &= \int_{x\in A}P(d|x,\hat{\theta},\alpha)\hat{F}(dx|A) \quad \text{(DP)}\end{aligned} \tag{6.1}$$

where $\hat{F}(dx|A)$ is the non-parametric estimate of the conditional probability distribution of x given A, $\hat{F}(dx|A) = N(dx)/N(A)$ where $N(x)$ is the number of observations in cell dx and $N(A)$ is the total number of observations in all x-cells in the set A.[23] Table 6.1 compares the NP and DP estimates of individuals' choice probabilities in the case where A is the universe of all possible x-cells. The DP model appears to fit the actual choice probabilities rather closely.

AB=apply for Social Security benefits, DB=don't apply

FT=full-time work, PT=part-time work, NE=not employed

Estimator	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	36.98	12.49	30.50	13.36	4.52	2.15
DP	37.55	12.75	29.86	12.79	4.25	2.80

Table 6.1 Predicted vs. actual choice probabilities: full sample

[23] Given a (possibly random) partition of the x cells, it is possible to compute an asymptotic Chi-squared goodness of fit test described by Andrews (1988). We do not present formal Chi-square goodness of fit tests here due to the fact that we found these statistics sensitive to the particular partition we chose for the x-cells. Also, Andrews' statistic requires consistent estimates of the asymptotic covariance matrix of the θ_u coefficients, but in view of our discussion of section 3 our estimates will only be consistent if the full information matrix is block diagonal. Since calculation of consistent covariance matrix estimates for θ_u is not computationally feasible, we did not think it would be informative to present Chi-square goodness of fit statistics under the maintained assumption of block diagonality. Instead we found it to be much more informative to directly compared predicted versus actual choice probabilities.

Since the RHS data follow individuals over the crucial period from age 58 to age 73 when they make the transition from full-time work to retirement, it makes sense to follow how the DP model tracks these transitions starting at age 58. Given that the DP model is based on biennial decision intervals, all of the following tables group individuals into two year age intervals. Thus, we begin our analysis by looking at the decisions of individuals aged 58-59. Individuals in this age range are not yet eligible for early retirement benefits at age 62, so their choice sets contain 3 alternatives: $\{FT, PT, NE\}$. Given that we have excluded individuals who have qualified for Social Security disability insurance, there are only 3 observations of men who are not working in this age group. Most of the men this age are working full-time, although there are 48 observations of men working part-time. Table 6.2 compares the predicted versus actual employment choices made by various subgroups of males aged 58-59 who were working full-time. The top panel of table 6.2 shows that the DP model predicts the behavior of the 705 individuals in this group quite closely, with approximately 94% deciding to continue working full-time and only 1% deciding to quit working. The second and third panels of table 6.2 illustrate how employment decisions vary by income class: individuals with low incomes and low average wages are correctly predicted to be nearly 3 times as likely to choose to quit working ($ed = NE$), whereas individuals with the highest incomes are correctly predicted to be more likely to continue working full-time and much less likely to take a part-time job or quit working. The bottom two panels of table 6.2 shows that the DP model correctly predicts the rather dramatic impact of health insurance coverage on employment decisions. The 563 individuals classified as having *eph* health insurance (i.e. health insurance through employer but no access to retiree health insurance) have a 95% chance of continuing to work full-time and less than one half of 1% chance of quitting work altogether whereas the remaining group of individuals who either have no health insurance or some kind of retiree health insurance or Medicaid have only an 88% chance of continuing to work full-time and a nearly 4% chance of quitting work. Given the relatively small numbers of observations in the bottom panel of table 6.2 we were unable to make statistically reliable comparisons for the *nhi, mca,* and *gph* groups separately. One might also question this finding on the grounds that the difference in the choice behavior in the last two panels of table 6.2 may actually be due to other factors, such as the 142 individuals in the last panel having lower income, poorer health, and being more likely to be single than the 563 individuals with *eph* insurance. However the DP model is able to predict the behavior impact of insurance coverage holding all of these other factors constant, and indeed we find that *cetaris paribus* individuals with *eph* health insurance coverage are significantly more likely to remain employed and are much less likely to quit working than any of the other groups. The explanation for this result is quite straightforward: individuals with *eph* insurance will lose their health insurance if they decide to quit work or work part-time. Since these individuals are risk averse and face a long Pareto-tailed distribution of health care liabilities, they have a much stronger incentive to remain employed full-time in order to retain their health insurance coverage than individuals who have retiree health insurance coverage or no health insurance.

$e = FT, N = 705$			
$\chi^2 = .718$	FT	PT	NE
NP	93.76	5.11	1.13
DP	94.43	4.65	0.92

$e = FT, y \leq 5, aw \in \{1, 2\}, N = 146$			
$\chi^2 = .454$	FT	PT	NE
NP	91.10	5.48	3.42
DP	90.67	6.53	2.80

$e = FT, y \geq 6, aw \in \{3, 4\}, N = 426$			
$\chi^2 = .410$	FT	PT	NE
NP	96.01	3.76	0.23
DP	96.41	3.45	0.14

$e = FT, hi = eph, N = 563$			
$\chi^2 = 3.650$	FT	PT	NE
NP	95.03	4.44	0.53
DP	96.50	3.15	0.35

$e = FT, hi \in \{gph, nhi, mca\}, N = 142$			
$\chi^2 = 1.267$	FT	PT	NE
NP	88.73	7.75	3.52
DP	86.23	10.62	3.15

Table 6.2 Predicted vs. actual choice probabilities for male full time workers aged 58-59

Table 6.3 compares predicted versus actual choice probabilities for the sample of 1214 men aged 60-61 who were working full-time. Note that the two-year decision interval implicit in our biennial specification of the DP model forces us to treat men at this age as making the decision as to whether or not they will submit an application for Social Security benefits in the following time period when they will be 62-63 and eligible for early retirement benefits. Thus, individuals in this age group are treated as having a choice set with 6 alternatives equal to the product of the three possible employment decisions *ed* $\in \{FT, PT, NE\}$ and the two possible Social Security application decisions, *ssd* $\in \{AB, DB\}$, where *AB* is a mnemonic for "apply for benefits" and *DB* is a mnemonic for "don't apply for benefits". The first panel of table 6.3 shows that the DP model is able to capture the age 62 peak in the retirement hazard function. We can see this by noting that the DP model predicts that 25.5% of eligible individuals apply for Social Security in this age range whereas in the data 23.9% apply. Also, note that the DP model predicts that the fraction of individuals choosing to quit work or work part-time jumps from 5.6% at ages 58-59 to 13.7% at age 60-61, compared to the NP estimates of 6.2% and 11.7%, respectively.[24]

Panels 2 through 4 of table 6.3 show how individuals' decisions are affected by the type of health insurance they have access to. Note that the DP model correctly predicts that individuals with *eph* health insurance are significantly more likely to continue working full-time and not apply for Social Security (77.7% NP, 73.8% DP) than individuals with *gph* insurance (44.2% NP, 55.2% DP), *mca* insurance (49.2% NP, 48.7% DP), or no health insurance, *nhi* (54.3% NP, 54.5% DP). Panels 6 and 7 of table 6.3 provide some perspective on the magnitude of the behavioral impact of health insurance by comparing how employment decisions are affected by being in bad health or by being single. It is well known that both of these groups are significantly less likely to continue working full time and significantly more likely to apply for Social Security than married men and those in good health. From Table 6.3 we can see that while these individuals are significantly less likely to continue working full-time and not apply for Social Security compared to the sample as a whole, the reduction in the probability of full-time work for these groups is not as great as for individuals with *gph, mca,* and *nhi:* the probability of working full-time is 88.3% for all men in table 6.3, compared to 91.5% for those with *eph,* 85.1% for those in bad health, 81.0% for those who are single, 76.8% for those with *gph,* 72.9% for those with *mca,* and 71.4% for those with *nhi.* The relative magnitudes of the effects of health, marital status, and health insurance on the probability of full-time work are not artifacts of correlated changes in other variables such as wages and income: the effects show up in predictions using the DP model holding these other variables constant. The DP model shows that health insurance has strong impacts on behavior, providing especially strong incentives for individuals with *eph* insurance to remain employed full-time even after they are eligible for Social Security early retirement benefits, whereas the other health insurance groups are significantly more likely to quit work or work part-time.

The next to last panel of table 6.3 shows that the DP model correctly predicts that high wage individuals are very likely to continue working full-time and not apply for early retirement benefits. The replacement rate for these individuals is low, so that retirement involves a substantial opportunity cost in terms of lost income. The final panel shows that the DP model correctly predicts that individuals who were working part-time at age 60-61 are much more likely to apply for Social Security benefits at age 62-63 (41.5% NP vs. 46.8% DP), and substantially less likely to return to a full-time job (48.9% NP vs. 44.8% DP). Overall, we conclude that the DP model is able to capture the peak in retirements at age 62, and the differential impacts of health, marital status, income, employment, and health insurance on individual behavior. In particular the DP model correctly predicts that any of the following individuals are significantly more likely to apply for Social Security early retirement benefits at age 62: those who are unhealthy,

[24] These figures were computed from the sum of columns 2 and 3 of table 6.1 and the sum of columns 2, 3, 5 and 6 of table 6.2.

$e = FT, N = 1214$						
$\chi^2 = 7.17$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	72.32	3.46	0.33	15.98	5.44	2.47
DP	70.01	3.67	0.79	16.31	6.00	3.21

$e = FT, hi = eph, N = 990$						
$\chi^2 = 14.93$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	77.68	2.83	0.30	13.84	4.75	0.61
DP	73.79	2.92	0.39	16.21	4.73	1.96

$e = FT, hi \in \{gph, mca, nhi\}, N = 224$						
$\chi^2 = 18.08$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	48.66	6.05	0.45	25.45	8.48	10.71
DP	53.27	7.01	2.59	16.75	11.63	8.74

$e = FT, h = B, N = 255$						
$\chi^2 = 4.58$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	63.53	2.75	0.39	21.57	7.06	4.71
DP	63.48	3.94	1.40	18.40	7.11	5.67

$e = FT, m = S, N = 137$						
$\chi^2 = 7.98$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	65.69	5.84	0.73	15.33	8.76	3.65
DP	58.26	3.83	2.60	21.62	8.19	5.55

$e = FT, y \geq 15, aw \geq 3, N = 30$						
$\chi^2 = 1.51$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	93.33	0.00	0.00	6.67	0.00	0.00
DP	95.33	1.25	0.00	3.26	0.12	0.05

$e = PT, N = 94$						
$\chi^2 = 6.41$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	39.36	19.15	0.00	9.57	25.53	6.38
DP	34.52	15.03	3.67	10.25	26.62	9.90

Table 6.3 Predicted vs. actual choice probabilities for men aged 60-61

single, those who have lower average wages, Medicaid recipients, individuals who have no health insurance, or individuals who have some form of retiree health insurance.

Table 6.4 compares the predicted versus actual choice probabilities for a sample of 1221 men aged 62-63 who haven't previously applied for Social Security. The DP model treats the decisions made at this age range as determining Social Security and labor supply status at ages 64-65. The first panel of table 6.4 shows that the DP model is able to capture the large jump in the "retirement hazard rate" at age 65: the fraction of individuals applying for Social Security increases from 25% of the age 62-63 group in table 6.3 to 68% of the age 64-65 age group in table 6.4. Specifically, the DP model predicts that 67.3% of individuals who haven't already applied for Social Security will apply at this age (compared to 67.7% in the data), and DP model predicts that 20.2% of these full-time workers will either stop working or switch to a part-time job (compared to 19.2% in the data). It is interesting to note the large percentage of men who apply for benefits yet continue working full-time (DP 49.7% vs. NP 49.9%). Most of these consist of the high wage, married individuals who continue working into their late 60's and early 70's. Table 6.4 also helps resolve a puzzle raised in section 2: why is it that virtually everyone applies for Social Security by age 65 even though a significant proportion of men clearly intend to continue working into their late 60's and early 70's? The DP model provides a relatively simple explanation for this phenomenon: it is optimal to apply for Social Security benefits at 65 even if one intends to continue working full-time beyond age 65 due to the *Medicare option:* applying at 65 amounts to exercising a costless option to supplement one's health care insurance with Medicare.[25]

The "Medicare option" is the key to the solution to the Medicare puzzle raised in section 3. Clearly the people who value the Medicare option the most are those with *eph* insurance: i.e. those wo do not have access to retiree health insurance. Comparing the second panel of table 6.3 with the second panel of table 6.4 we see a significant shift in labor supply behavior between ages 62-63 and 64-65: nearly 92% of the *eph* individuals decide to continue working full-time at age 62-63 (85% of which decide not to apply for Social Security) whereas at age 64-65 only 83% decide to continue working full-time (only 38% of which decide not to apply for Social Security). Panels 3 to 5 of table 6.4 show that individuals with *gph* and *mca* health insurance or those with no health insurance have only a 70% chance of continuing to work full-time. The explanation for this difference in behavior is clear: by continuing to work full-time, the *eph* individuals are essentially obtaining dual coverage from their employer's health plan and Medicare. As we showed in section 4.4 this dual coverage significantly reduces the risks of incurring catastrophic medical bills compared to Medicare alone. Thus, it appears that many of the *eph* individuals are continuing to remain employed full-time after

[25] Note that the provision for automatic recomputation of benefits implies that there is no sacrifice in terms of computed AME for applying at 65 and continuing to work on a full-time basis: all earnings made after age 65 will be credited to the individual's account until the first month in which he starts actually collecting benefits. It is true that if these individuals continue to work after applying for Social Security they will lose all of their benefits to the Social Security earnings tax, but it is incorrect to view this as a "cost" to applying at 65. If the individual hadn't applied for Social Security at 65 he would be in exactly the same position as if he had applied with the exception of being without Medicare coverage. In this sense, application for Social Security benefits at age 65 amounts to costlessly exercising an option to obtain Medicare coverage.

$e = FT, N = 1221$						
$\chi^2 = 20.57$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	30.96	1.06	0.25	49.88	14.09	3.77
DP	30.15	1.89	0.63	49.70	11.91	5.73

$e = FT, hi = eph, N = 1065$						
$\chi^2 = 29.22$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	31.46	0.94	0.09	50.99	14.08	2.44
DP	31.12	1.68	0.41	50.79	11.00	5.00

$e = FT, hi \in \{gph, mca, nhi\}, N = 156$						
$\chi^2 = 4.64$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	27.56	1.92	1.28	42.31	14.10	12.82
DP	23.46	3.31	2.16	42.21	18.14	10.71

$e = FT, h = B, N = 270$						
$\chi^2 = 11.01$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	26.67	1.48	0.74	48.52	17.41	5.19
DP	28.14	2.03	1.04	46.84	12.59	9.36

$e = FT, m = S, N = 130$						
$\chi^2 = 13.77$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	38.46	0.00	0.77	43.85	11.54	5.38
DP	26.25	1.78	1.40	45.38	16.11	9.08

$e = FT, y \geq 15, aw \geq 3, N = 30$						
$\chi^2 = 1.44$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	30.32	0.00	0.00	60.47	6.98	2.33
DP	33.88	1.38	0.16	53.19	8.22	3.17

$e = PT, N = 75$						
$\chi^2 = 8.11$	FT,DB	PT,DB	NE,DB	FT,AB	PT,AB	NE,AB
NP	17.33	9.33	2.67	32.00	29.33	9.33
DP	13.12	6.70	2.89	22.74	41.85	12.70

Table 6.4 Predicted vs. actual choice probabilities for men aged 62-63 not already receiving Social Security

applying for Medicare in order to retain this "deluxe" health care coverage. On the other hand, individuals with *gph* health insurance do not need to remain employed in order to obtain the deluxe combined private/Medicare coverage, and this is reflected in their lower likelihood of remaining on their full-time job. Similar effects are observed for individuals with Medicaid or those without private health insurance, although the relatively small numbers in these latter groups implies that the NP estimates of their response are quite noisy. However all these effects have been verified in simulations of the DP model holding income and the values of other state variables constant. Overall, we conclude that modeling the incompleteness of health insurance markets is critical to understanding differential responses to Social Security incentives. Even though health care enters the model in a very subtle and indirect way (i.e. by shifting the subjective distribution of total family income), it has a major impact on the ability of the DP model to fit the data, increasing the likelihood function by over 100 points.

The remaining four panels of table 6.4 show the impact of health, marital status income, and employment status on individuals' decisions. Compared to the full-sample of full-time workers in panel 1 of table 6.4, we see that single persons and individuals who are in bad health have more than a 50% greater chance of quitting work, whereas high wage individuals are significantly more likely to continue working full-time (90.8% versus 80.8% for the full sample). However perhaps the most important factor affecting individuals' decisions is their current employment status: the last panel of table 6.4 shows that if the individual is currently working part-time, they have less than a 50% chance of returning to full-time work.

We conclude our analysis with table 6.5 which summarizes the strong disincentive effects of Social Security and Medicare on labor supply. The first two panels of table 6.5 compare the labor supply decisions of the sub-population of individuals over 64 who are working full-time and who have already applied for Social Security ($ss \in \{ER, NR\}$) versus those who are not eligible for benefits since they haven't yet applied ($ss = NE$). We see that the DP model correctly predicts the large incentive effects on labor supply decisions: full-time workers who are receiving Social Security benefits are significantly less likely to continue working full-time, are somewhat more likely to choose to work part-time, and are more than twice as likely to decide to stop working altogether than individuals who are not receiving Social Security benefits. Panels 3 and 4 make the same comparison for the larger sub-population of individuals who are working on a full or part-time basis. Here the difference in labor supply behavior is even larger, with individuals who are not receiving Social Security being nearly twice as likely to choose to work full-time than individuals receiving Social Security. The final two panels compare the full population of all men over 65. We obtain the most dramatic difference in behavior for the full population: individuals who are not receiving Social Security benefits being more than three times as likely to work full-time and only one third as likely to stop working than those receiving Social Security benefits. The shift is due the fact that "full retirement" (i.e. when an individual stops working *and* begins receiving Social Security benefits) is nearly an absorbing state: the DP model correctly predicts that over 90% of the fully retired men in our sample choose to remain out of the labor force: only 1% choose to return to full time work,

$e = FT, t \geq 65, ss \in \{ER, NR\}, N = 1077$			
$\chi^2 = 0.70$	FT	PT	NE
NP	52.55	22.19	25.26
DP	52.61	23.05	24.34

$e = FT, t \geq 65, ss = NE, N = 289$			
$\chi^2 = .10$	FT	PT	NE
NP	70.69	16.26	13.15
DP	69.41	17.09	13.50

$e \in \{FT, PT\}, t \geq 65, ss \in \{ER, NR\}, N = 1767$			
$\chi^2 = .47$	FT	PT	NE
NP	36.67	34.47	28.86
DP	37.43	34.23	28.34

$e \in \{FT, PT\}, t \geq 65, ss = NE, N = 312$			
$\chi^2 = .08$	FT	PT	NE
NP	66.67	18.91	14.42
DP	65.92	19.50	14.58

$e \in \{FT, PT, NE\}, t \geq 65, ss \in \{ER, NR\}, N = 3445$			
$\chi^2 = .74$	FT	PT	NE
NP	19.30	20.00	60.70
DP	19.81	20.16	60.03

$e \in \{FT, PT\}, t \geq 65, ss = NE, N = 334$			
$\chi^2 = .18$	FT	PT	NE
NP	62.28	18.87	18.86
DP	61.70	18.53	19.77

Table 6.5 Summary of the incentive effects of Social Security and Medicare on male labor supply

and approximately 5% choose part-time work. Much of the disincentive effect can be ascribed to the Social Security earnings test, although some of the disincentive is also due to the high psychic “search costs” of labor market reentry. [26]

[26] We acknowledge that the high estimated search costs may also reflect the effects of age discrimination, which suggests that it may be inapropriate to treat them as “structural parameters” characterizing individual preferences. While it is possible to estimate “imperfect control” versions of the DP model that allow for a positive probability that an older unemployed worker searching for a job will be unsuccessful (or an employed worker will be

As we discussed in section 3, once the DP model is estimated, it is a simple matter to perform counterfactual predictions of the impact of alternative Social Security policies: simply re-solve the DP model with a new Social Security policy α' instead of the historical policy α. A particularly dramatic forecasting exercise is to predict what individual behavior would be like in the absence of Social Security. When we do this, we find that the new optimal decision rule without Social Security implies much higher rates of labor force participation and the peaks in the hazard rates for labor force exits at ages 62 and 65 *disappear completely,* just as they did in policy simulations of Gustman and Steinmeier's (1986) structural model. These results suggest that many of the features of observed retirement behavior are indeed artifacts of the details of the U.S. Social Security rules.[27]

7. Conclusion

We are now in a position to answer the question raised in the title to this paper: how do Social Security and Medicare affect retirement behavior in a world of incomplete markets? Our results suggest that Social Security creates significant disincentives to labor force participation, and is "responsible" for the peaks in retirements at ages 62 and 65, the ages of eligibility for early and normal Social Security retirement benefits. Our approach to inferring the behavioral impacts of Social Security and Medicare involved constructing a relatively simple and parsimonius model of how rational agents optimally respond to the rules in effect during the 1970's and demonstrating that the predictions of the DP model closely mimic the actual behavior of a sample of men whose only pension plan is Social Security. The DP model allows us to conduct the conceptual experiment of comparing the behavior of two otherwise identical individuals, one of whom is entitled to Social Security benefits. We found that an individual who is employed and entitled to Social Security benefits is significantly less likely to continue working than his counterpart who is not entitled. Conversely an unemployed individual who is entitled to Social Security has very little chance of returning to work in comparison to an identical individual who is not entitled. The most radical experiment — eliminating Social Security entirely — ends up completely eliminating the peaks in retirements at ages 62 and 65.

We believe that accurate modeling of the Social Security rules and the the nature of market incompleteness is necessary in order to capture the range of individual behavioral responses to these rules, which differ markedly as a

laid off), it is very difficult to separately identify parameters characterizing the effects of age discrimination, psychic search costs, and the disutility of working. In fact it is a very challenging measurement problem to simply distinguish individuals who were unsuccessful in job search or who were involuntarily unemployed from those who are not working by choice due to the common tendency of older individuals to report employment status as "retired" if not working. The perfect control assumption is rendered more palatable by recalling that although we assume an individual can find *some* job with probability 1, there is no guarantee that it will be a "good" job (i.e. a high wage job). Furthermore the estimation results in section 4.2 already embody a form of "age discrimination" in as much as the distribution of wages drifts downward as the individual ages.

[27] We do not present the specific output of this counterfactual simulation since its only purpose is to emphasize the strong incentive effects of Social Security that have already been amply illustrated in the many tables presented above. We don't take the DP model's predictions of the impact of such radical changes in policy seriously because our model does not account for general equilibrium repercussions to an end to Social Security, including changes in wage profiles, in individual saving behavior, and the presumable expansion of private markets for annuities and health insurance. In future work we will use the DP model to predict the effects of less dramatic changes in policy (such as the effects of the 1983 Social Security amendments) and use this as a basis for out-of-sample predictive tests of the model.

function of age and individual characteristics. Accounting for three particular types of market incompleteness seems essential to an accurate account of how Social Security and Medicare affect individual decision-making: 1) incomplete annuities markets, 2) borrowing constraints, and 3) incomplete markets for health insurance. Incompleteness in the annuities markets is necessary condition for Social Security to have a significant behavioral effect, since if individuals had access to a rich market of fairly priced annuities they would be able to design their own personal optimal retirement plans rather than relying on Social Security, using offsetting private transfers to neutralize the distortionary effects of the progressive tax/transfer features of Social Security.[28] The DP model accounts for incomplete annuity markets by assuming that individuals' only annuity plan is Social Security. We appealed to another form of incomplete markets — borrowing constraints — to explain the existence of the peak in retirements at age 62. The DP model accounts for borrowing constraints via the assumption that consumption equals income period by period. While we have not proven that such a decision rule is optimal, we have argued that it is a good approximation for the men in our sample, since most have accumulate very little non-housing net worth available to finance any significant amount of pre-retirement consumption. A final form of market incompleteness — incomplete markets for private health insurance — appears to be a key part of the explanation for the peak in retirements at age 65. We have accounted for incomplete health insurance markets by assuming that certain individuals only have access to private health insurance via their employer's group health plan but have no access to retiree health benefits (*eph*), whereas others have access to fairly priced health insurance independent of whether or not they are employed (*gph* and *mca*), and that still others have no access to any form of private health insurance whether or not they are employed (*nhi*). Based on data available in the RHS, we have found that most individuals are in the first class, *eph*, and these individuals have a strong incentive to remain employed until they are eligible for Medicare insurance at age 65. The DP model correctly predicts that the other individuals with *gph, mca,* or *nhi* health insurance are significantly more likely to apply for Social Security early retirement benefits at age 62.

This paper has succeeded in providing answers to two of the three empirical puzzles raised in the introduction, namely, the "age 65 retirement puzzle" and the "Medicare puzzle". With the exception of the model by Gustman and Steinmeier (1986), most previous econometric studies have had difficulty accounting for the large peak in retirements at age 65. In addition previous structural analyses of the impact of Medicare (Gustman and Steinmeier 1993 and Lumsdaine, Stock and Wise, 1994, 1995) have concluded that Medicare has virtually no impact on labor supply, a result that conflicts with the findings of reduced-form studies and the conventional wisdom of the "man of the street". Instead of valuing Medicare as the expected value of Medicare reimbursements, the DP model explicitly accounts for the long thin Pareto-tailed distribution of health care costs. Even though the expected value of Medicare benefits is small, our structural parameter estimates indicate substantial risk aversion on the part of individuals which leads

[28] See Crawford and Lilien (1981). Bernheim and Bagwell (1988) provide an interesting discussion of much more general policy neutrality result, when both inter and intra-generational transfers are taken into account.

the DP model to assign a very high certainty equivalent valuation to the "Medicare option" — especially for the *eph* individuals who do not have access to retiree health benefits. The other pieces of the age 65 retirement puzzle have already been recognized in the previous literature: the negligible 1% delayed retirement credit implies that Social Security is substantially actuarially unfair after age 65, and the Social Security earnings test leads to 50% tax surcharge on wage earnings above a small threshold level, providing significant disincentives to continued labor force participation. The DP model shows how the interaction of all of these effects is responsible for many of the features of observed retirement behavior such as the fact that virtually everyone applies for Social Security on or before their 65^{th} birthday, and the fact that the peak in applications for Social Security benefits at age 65 also corresponds to a peak in labor force exits at that age as well.

We conclude by pointing out a number of limitations of our approach. First, since our paper focuses on a subset of men whose only pension plan is Social Security, our model cannot address the "early retirement puzzle", i.e. the reasons behind the historical decline in aggregate male labor force participation rates. Our results suggest that Social Security has been a key element in this decline, but until we generalize the DP to incorporate disability insurance and private pensions will we be unable to determine the relative importance of Social Security, private pensions, and disability insurance as potential explanations for the acceleration in the decline in labor force participation during the 1970's. Second, because we have assumed that $c_t = y_t$ (which we argued is empirically justified and may be consistent with optimal behavior in the presence of borrowing constraints) our model is incapable of predicting how people would respond to Social Security in a world of *complete markets* where in particular there are no borrowing constraints. Although theory tells us that Social Security policy should have smaller effects on retirement behavior in such a world, the distortionary features of Social Security (e.g. the earnings test tax and the actuarial unfairness of the benefit schedule) will still lead to a peak in retirements at age 65, although it is hard to see how Social Security *per se* could produce a peak at age 62 if borrowing constraints did not exist. We think it is especially important to relax the $c_t = y_t$ assumption in order to understand behavior at earlier stages in the life-cycle and to address the "consumption-savings puzzle" i.e. why these men failed to accumulate significant pre-retirement net worth. One could argue that low savings rates are a rational response to high anticipated Social Security benefit levels, especially for the RHS cohorts where the present value of Social Security benefits greatly exceeded the present value of Social Security taxes. However it is difficult to use this same argument to explain the low savings rates of later generations (especially the "baby boomers" for whom the present value of Social Security taxes are likely to greatly exceed the present value of Social Security benefits). It remains to be seen whether the trend towards early retirement will eventually be reversed, and whether low rates of private savings can be "rationalized" by expectations of delayed age of retirement. However we believe that by extending the basic DP framework of this paper to account for pensions, disability, and consumption-savings decisions we may eventually be able to shed light on these puzzles.

8. Appendix: construction and validation of health expenditure and health insurance classification variables

This appendix describes the procedures used to classify individuals into one of the 4 possible health insurance states eph, gph, mca or nhi defined in section 3 using the RHS data. The RHS asked respondents whether they are covered by Medicare or Medicaid, whether they have any type of private health insurance such as Blue Cross Blue Shield, and whether they have any health insurance paid entirely by their employer or union. However the survey did not ask individuals who were not covered by private health insurance whether they had access to retiree health insurance, or had ever been turned down in an application for insurance. Without such information it is difficult to determine whether uninsured individuals are uninsured by choice or uninsured due to market incompleteness. Furthermore we don't have sufficient information to tell whether individuals who are currently covered by an employee health policy have the option of converting their coverage into a retiree health plan after they leave the firm. In view of the data limitations, we resorted to the following relatively crude classification rule: we categorized an individual as having gph, (i.e. general private health insurance such as Blue Cross which is not tied to employment) if we ever observed that the individual was unemployed and reporting that he was covered by a private health plan of some type. We classified someone as being in state nhi (no health insurance) if they never reported being covered by a private health plan in any wave of the survey. We classified someone as being in state mca (Medicaid recipient) if they reported receiving Medicaid in some wave of the RHS, and the remaining individuals were classified as being in state eph, i.e. as having employer-provided health insurance. Our presumption is that eph individuals do not have access to retiree health insurance if they are not employed: otherwise we would have observed the respondent reporting that they had private health when they weren't working and would have already classified them as gph. Approximately 56% of the sample were classified as having eph insurance. This is undoubtedly an overestimate: about 22.5% of the individuals in our sample reported working full time in all waves of the survey, all of whom are classified as eph by this procedure. We have verified the robustness of our results by focusing on a subset of *eph* individuals who we observed both while they were working and while they were not working. A total of 1400 such individuals reported having private health insurance coverage while they were working and no form of private health insurance whenever they were not working. It seems to be safe to conclude that these 1400 individuals did not have access to fairly priced private health insurance, otherwise they would have purchased it and would have already been classified as gph under our procedure. When we plot the distribution of the age of first receipt of OADI benefits for this subsample of "verified" *eph* individuals, we find it is not substantially different: the peak at 62 increases by 5 percentage points and the peak at 65 decreases by 5 percentage points.

9. References

Andrews, D.W.K. (1988) “Chi-Square Diagnostic Tests for Econometric Models” *Journal of Econometrics* **37** 135–156.

Berkovec, J. and S. Stern (1991) “Job Exit Behavior of Older Men” *Econometrica* **59-1** 189–210.

Bernheim, B.D. (1988a) “Social Security Benefits: A Study of Expectations and Realizations” in Rita Ricardo-Cambell and E. P. Lazear (eds.) *Issues in Contemporary Retirement* Hoover Institution, Stanford, California.

Bernheim, B.D. (1988b) “How Do the Elderly Form Expectations? An Analysis of Responses to New Information” NBER Working Paper 2719.

Bernheim, B.D. and K. Bagwell (1988) “Is Everything Neutral?” *Journal of Political Economy* **96-2** 308–338.

Blau, D. (1994) “A Structual Model of Retirement-Related Labor Market Dynamics” *Econometrica* **62-1** 117–156.

Blinder, A.S. Gordon R.H. and Donald E. Wise (1980) “Reconsidering the Work Disincentive Effects of Social Security” *National Tax Journal* **33** 431–442.

Boskin, M. (1977) “Social Security and Retirement Decisions” *Economic Inquiry* **15**, 1–25.

Boskin, M. and M. Hurd (1978) “The Effect of Social Security on Early Retirement” *Journal of Public Economics* **10** 361–377.

Bound, J. (1989) “The Health and Earnings of Rejected Disability Insurance Applicants” *American Economic Review* **79** 482–503.

Bound, J. (1991) “Self-Reported Health Vs. Objective Measures of Health in Retirement Models” *Journal of Human Resources* **26** 106–138.

Burkhauser, R.V. and Turner, J.A. (1981) “Can Twenty-Five Million Americans Be Wrong? A Response to Blinder, Gordon and Wise” *National Tax Journal* **34** 467–472.

Burtless, G. (1986) “Social Security, Unanticipated Benefit Increases, and the Timing of Retirement” *Review of Economic Studies* **53** 781–805.

Burtless, G. and Moffitt, R. (1984) “The Effect of Social Security Benefits on the Labor Supply of the Aged” in H. Aaron and G. Burtless (eds.) *Retirement and Economic Behavior* Brookings Institution, Washington, D.C.

Cochrane, J. (1989) “The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives” *American Economic Review* **79-3** 319–337.

Crawford, V. and Lilien, D. (1981) “Social Security and the Retirement Decision” *Quarterly Journal of Economics* **3** 505–534.

Deaton, A. (1991) “Saving and Liquidity Constraints” *Econometrica* **59-5** 1221–1248.

Fields, G.S. and O.S. Mitchell (1984) *Retirement, Pensions, and Social Security* MIT Press, Cambridge, Massachusetts.

Friedman, B. and M. Warshawsky (1988) “Annuity Prices and Savings Behavior in the United States” in Z. Bodie, J. Shoven and D.A. Wise (eds.) *Pensions in the U.S. Economy* University of Chicago Press.

Geweke, J. Zarkin, G. Slonim, R. Hiedemann, B. and P. Atkison (1992) “The Social Security Acceptance Decision”, manuscript Duke University.

Gordon, R. and A.S. Blinder (1980) “Market Wages, Reservation Wages, and the Retirement Decision” *Journal of Public Economics* **14** 277–308.

Gruber, J. and B.C. Madrian (1993) “Health Insurance and Early Retirement: Evidence from the Availability of Continuation Coverage” NBER Working Paper 4594.

Gustman, T. and T.L. Steinmeier (1986) “A Structural Retirement Model” *Econometica* **54-3** 555-584.

Gustman, T. and T.L. Steinmeier (1985) “The 1983 Social Security Reforms and Labor Supply Adjustments of Older Individuals in the Long Run” *Journal of Labor Economics* **3** 237–253.

Gustman, T. and Steinmeier, T.L. (1993) “Employer Provided Health Insurance and Retirement Behavior” NBER Working Paper 4307.

Gustman, T. and Steinmeier, T.L. (1993) “Retirement in a Family Context: A Structural Model for Husbands and Wives” manuscript, Dartmouth College.

Hurd, M.D. (1990) “Research on the Elderly: Economic Status, Retirement, and Consumption and Saving” *Journal of Economic Literature* **28** 565–637.

Hurd, M.D. (1993) “The Effect of Labor Market Rigidities on the Labor Force Behavior of Older Workers” manuscript, State University of New York at Stony Brook.

Kahn, J.A. (1988) “Social Security, Liquidity, and Early Retirement” *Journal of Public Economics* **35** 97–117.

Keane, M. P. and K.I. Wolpin (1994) “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation: Monte Carlo Evidence” manuscript, New York University.

Krueger, A. and J. Pischke (1991) “The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation” NBER working paper 3699.

Lumsdaine, R. and D.A. Wise (1994) “Aging and Labor Force Participation: A Review of Trends and Explanations” in D.A. Wise (ed.) *Aging in the United States and Japan: Economic Trends* University of Chicago Press.

Lumsdaine, R. Stock, J. and D.A. Wise (1992) “Three Models of Retirement: Computational Complexity vs. Predictive Validity” in D.A. Wise (ed.) *Topics in the Economics of Aging* University of Chicago Press.

Lumsdaine, R. Stock, J. and D.A. Wise (1994) “Pension Plan Provisions and Retirement: Men & Women, Medicare, and Models” in D.A. Wise (ed.) *Studies in the Economics of Aging* University of Chicago Press.

Lumsdaine, R. Stock, J. and D.A. Wise (1995) “Why Are Retirement Rates So High at Age 65?” NBER working paper 5190.

Lundberg, S. (1985) “Tied Wage-Hours Offers and the Endogeneity of Wages” *Review of Economics and Statistics* **66** 405–410.

MaCurdy, T.E. (1983) “A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty” *International Economic Review* **24-2** 265–289.

Moffitt, R.A. (1987) “Life-Cycle Labor Supply and Social Security: A Time-Series Analysis” in G. Burtless (ed.) *Work, Health, and Income among the Elderly* The Brookings Institution, Washington, D.C.

Myers, R.J. (1982) “Why Do People Retire Early?” *Aging and Work* **5-2** 83–91.

Myers, R.J. (1983) “Further About Controversy on Early Retirement Study” *Aging and Work* **6-2** 105–109.

Phelan, C. and Rust, J. (1993) “U.S. Social Security Policy: A Dynamic Analysis of Incentives and Self-Selection” manuscript University of Wisconsin.

Quinn, J.F. Burkhauser, R.V. and D.A. Myers (1990) *Passing the Torch: The Influence of Economic Incentives on Work and Retirement* W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.

Ransom, R.L. and R. Sutch (1988) "The Decline of Retirement in the Years Before Social Security: U.S. Retirement Patterns 1870-1937" in R. Ricardo-Campbell and E. Lazear (eds.) *Issues in Contemporary Retirement* Hoover Institution Press, Stanford, California.

Ruhm, C.J. (1990) "Bridge Jobs and Partial Retirement" *Journal of Labor Economics* **8-4** 482–501.

Rust, J. (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher" *Econometrica* **55-5** 999–1033.

Rust, J. (1988) "Maximum Likelihood Estimation of Discrete Control Processes" *SIAM Journal on Control and Optimization* **26-5** 1006–1023.

Rust, J. (1989) "A Dynamic Programming Model of Retirement Behavior" in D.A. Wise (ed.) *The Economics of Aging* University of Chicago Press, 359–398.

Rust, J. (1990) "Behavior of Male Workers at the End of the Life-Cycle: An Empirical Analysis of States and Controls" in D.A. Wise (ed.) *Issues in the Economics of Aging* University of Chicago Press.

Rust, J. (1992) "A Dynamic Programming Model of Retirement Behavior" in K.R. Billingsley, H.U. Brown III, and E. Derohanes (eds.) *Computer Assisted Analysis and Modelling on the IBM3090 The 1989 IBM Contest Prize Papers* Baldwin Press, volume 2, 887–912.

Rust, J. (1993) "Survey Design of the RHS and the HRS from A Decision-Theoretic Perspective" manuscript prepared for the HRS Early Results Workshop, Ann Arbor.

Rust, J. (1994a) "Estimation of Dynamic Structural Models: Problems and Prospects Part I: Discrete Decision Processes" forthcoming in C. Sims and J.J. Laffont (eds.) *Proceedings of the 6th World Congress of the Econometric Society* Barcelona, Spain, Cambridge University Press.

Rust, J. (1994b) "Structural Estimation of Markov Decision Processes" forthcoming in R. Engle and D. McFadden (eds.) *Handbook of Econometrics* volume 4, North Holland, Amsterdam.

Rust, J. (1995) "Using Randomization to Break Bellman's 'Curse of Dimensionality' " manuscript, University of Wisconsin (also in econwpa.wustl.edu under "Computational Economics").

Rust, J. (1996) *Dynamic Programming and Retirement Behavior* manuscript, University of Wisconsin.

Stewart, J. (1995) "Do Older Workers Respond to Changes in Social Security Benefits? A Look at the Time Series Evidence" manuscript, U.S. Bureau of Labor Statistics, Washington, D.C.

Stock, J. and D.A. Wise (1990) "Pensions, The Option Value of Work and Retirement" *Econometrica* **58–5**, 1151–1180.

Venti, S.F. and D.A. Wise (1990) "But They Don't Want to Reduce Housing Equity" in D.A. Wise (ed.) *Issues in the Economics of Aging* University of Chicago Press.