Volume Title: Issues in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press, 1990

Volume ISBN: 0-226-90297-8

Volume URL: http://www.nber.org/books/wise90-1

Conference Date: May 19-21, 1988

Publication Date: January 1990

Chapter Title: Behavior of Male Workers at the End of the Life Cycle: An Empirical Analysis of States and Controls

Chapter Author: John P. Rust

Chapter URL: http://www.nber.org/chapters/c7122

Chapter pages in book: (p. 317 - 382)

11 Behavior of Male Workers at the End of the Life Cycle: An Empirical Analysis of States and Controls

John Rust

This is the second installment in a series of three papers studying the behavior of men at the end of the life cycle. The first paper (Rust 1989) constructed a theoretical model based on the hypothesis that workers maximize expected discounted lifetime utility. The model treats observed behavior as a realization of a *controlled stochastic process* $\{x_t, d_t\}$ derived from the solution to a stochastic dynamic programming problem (DP). Estimation of the DP model requires observations of the worker's *state* x_t and *control* d_t and a specification of the Markov transition probability density $\pi(x_{t+1}|x_t, d_t)$ representing a stochastic "law of motion" that embodies workers' beliefs of uncertain future events.

This paper uses the Retirement History Survey (RHS) to construct state and control variables $\{x_{ti}, d_{ti}\}$, $t = 1, \ldots, T_i$, $i = 1, \ldots, I$, for a sample of $I = 8{,}131$ male respondents interviewed biennially from 1969 to 1979. I discuss some of the conceptual problems involved in constructing measurements of $\{x_t, d_t\}$ so that the resulting discrete-time, discrete-state DP model makes the best possible approximation to the underlying continuous-time, continuous-state decision process. I present my solutions to the measurement problems and conduct an extensive comparative data analysis to assess the

John Rust is professor of economics and director of the Data and Computation Center and the Social Systems Research Institute at the University of Wisconsin and a research associate of the National Bureau of Economic Research.

The author is extremely grateful for financial support from the National Institute on Aging and the National Science Foundation, which provided a work station for preparing the data and supercomputer time for estimating the DP model. He is equally thankful to the Institute for Empirical Macroeconomics at the Federal Reserve Bank of Minneapolis, which financed six months of uninterrupted research time during which this research was completed. Finally, he would like to thank B. J. Lee for his able research assistance and Lawrence Christiano, Finn Kydland, Edward Prescott. David Runkle, Christopher Sims, and Dick Todd for helpful comments on his work during seminars presented at the Federal Reserve.

overall quality of the resulting variables. Finally, I present estimates of workers' beliefs, in the form of an estimated transition probability matrix $\hat{\pi}$.

All this work is building up to the third paper of the series, which will use the constructed state and control variables and the estimated transition probability matrix as inputs to a "nested fixed point" maximum likelihood algorithm (Rust 1988) to estimate the unknown parameters of workers' utility functions. The success of the final stage depends critically on accurate measurements of $\{x_t, d_t\}$ and correct specification of workers' beliefs π.

The paper is organized as follows. The key parts of the paper are sections 11.1 and 11.2, which summarize the principal findings. Section 11.1 describes the state and control variables constructed from the RHS data set and presents the main conclusions of the data analysis. Section 11.2 specifies the functional form of workers' beliefs and summarizes the conclusions about workers' beliefs π. The remaining sections present details on the construction of $\{x_t, d_t\}$ and the specification and estimation of $\hat{\pi}$ that compose the evidence for the conclusions drawn in sections 11.1 and 11.2; they can be skipped or skimmed by readers who are content to accept my view of the "stylized facts."

11.1 State and Control Variables: Main Findings

Following the notation of Rust (1989), the DP model requires a vector of state variables, $x_t \equiv (w_t, y_t, aw_t, sr_t, h_t, a_t, e_t, ms_t)$, defined by

w_t: accumulated net financial and tangible nonfinancial wealth,
y_t: total income from earnings and assets,
aw_t: Social Security "average monthly wage,"
sr_t: Social Security status (receiving OASDI/not receiving OASDI),
h_t: health status of worker (good health/poor health/disabled/dead),
a_t: age of worker,
e_t: employment status (full time/part time/not employed),
ms_t: marital status (married/single),

and control variables, $d_t \equiv (c_t, s_t, ss_t)$, defined by

c_t: planned consumption expenditures,
s_t: employment search decision (full time/part time/exit labor force),
ss_t: Social Security decision (apply for OASDI/do not apply for OASDI).

In the last twenty years, several panel data sets have accumulated sufficiently detailed data to permit construction of the required variables: the Panel Survey on Income Dynamics (PSID), the National Longitudinal Survey (NLS), and the Retirement History Survey (RHS). Of these, the RHS has the largest and most comprehensive coverage of older workers since it was explicitly designed by the Social Security Administration (SSA) to obtain a detailed picture of the transition from work into retirement. A special feature of the RHS is the availability of matching records from the Census Bureau and

SSA that permit direct validation of response error in several key variables. The Social Security Earnings Record (SSER) contains each covered worker's wage earnings (up to the statutory maximum taxable earnings) and quarters of coverage from 1939 to 1974. The Social Security Master Beneficiary Record (SSMBR) contains actual payments of Social Security old age, survivors, disability, and death benefits (OASDI) to each respondent, spouse, and dependent from 1969 to 1978. The combination of finely detailed data, large sample size, and long duration, plus the existence of linked Census and SSA records, makes the RHS the data set of choice for estimating the DP model.

Having said this is not to deny the sober truth that, even with the linked RHS records, there is a limit to how accurately one can measure the "true" states and decisions of individuals. Besides the obvious problems of missing data, response and coding error, estimation of the DP model presents three additional problems: choice of time discretization, choice of state discretization, and construction of observable indicators of latent state and control variables.

Although the individual's actual decision process is best modeled in continuous time, the data are collected and the DP model is formulated in discrete time. In theory, use of discrete-time models is not a limitation since it has been shown that under very general conditions one can formulate a discrete-time DP model that approximates an underlying continuous-time DP model arbitrarily closely as the time interval goes to zero (van Dijk 1984). One can account for absence of data on (x_t, d_t) between survey dates by forming a marginal likelihood function that "integrates out" the missing observations. In practice, however, computational and data limitations forced me to use fairly coarse two-year time intervals. The computational limitations arise from the numerical integrations required to form the marginal likelihood function and the "curse of dimemsionality" inherent in DP models with fine time grids. The data limitations stem from the lack of measures of income flows for intervals shorter than one year.

Even if a complete set of "instantaneous" state and control variables could be constructed, I would still prefer to use annual measures in the belief that they better capture the worker's retirement behavior than a series of "snapshots" at fairly widely spaced time intervals. Analytically, the disadvantage of the discretization is that it implicitly precommits workers to fixed consumption and labor supply values over two-year time intervals. I should point out, however, that, even with a two-year time interval, the worker is given thirty opportunities to revise his decisions between age 58 and the terminal age, 108. Such a model is quite a bit more flexible than the standard approach, which models retirement as a once and for all choice from a nonlinear budget set that describes alternative consumption/work levels that are assumed fixed for the duration of the worker's lifetime (see, e.g., Burtless and Moffitt 1984). These sorts of static perfect-certainty models do not allow for any *ex post* revision in consumption or labor supply in light of new information. Whether a model with thirty periods (sixty years) can provide sufficient flexibility to model workers' decision processes accurately is a deeper question. I leave the analysis of the consequences of time aggregation to the actual estimation of the DP model in the third paper of this series.

The state variables y_t, w_t, and aw_t, which are typically treated as continuous, must also be discretized in order to estimate the DP model. Similar to the discretization of time, there are theorems guaranteeing that one can approximate a continuous-state DP model arbitrarily closely by a discrete-state DP model (Bertsekas and Shreve 1978). In previous work, I have found that the DP solution is not very sensitive to the discretization of the state variables and that one can obtain a good approximation using fairly coarse grids (Rust 1987). In this study, I use a grid size of \$1,000 (1968 dollars), which turned out to be more than adequate given the two-year time discretization that I ultimately adopted.

The most difficult problem, however, was construction of good measurements of latent variables such as health h_t, labor search s_t, and consumption c_t. My approach was to use all relevant survey responses to define observable indicators that might be regarded as "best approximations" of the underlying latent variables. Measurement of consumption proved to be particularly challenging, a fact that many economists may find disturbing. Even though the RHS asked respondents to list the amount spent on individual consumption items, in my opinion the list was too incomplete to construct reliable estimates of total consumption.[1] Since the RHS has much more complete, detailed data on income and wealth,[2] my approach was to infer c_t from the budget equation

$$(1) \qquad w_{t+1} = w_t + y_t - c_t \, .$$

Unfortunately, the RHS recorded income only in the even-numbered years immediately preceding each survey date. Thus, in order to construct c_t I needed to impute income in odd-numbered years. This in turn necessitated construction of complete labor force histories for each worker, including total annual hours worked in each year.[3] Using hours worked together with annual wage earnings data from the SSER (available up to 1974), I was able to impute income in odd-numbered years and construct estimates of c_t over the two-year sample interval. A limitation of the income data is absence of information on capital gains. I dealt with this problem by attributing 100 percent of the change in house value to capital gains (provided the respondent was a homeowner and had not moved within the interval) and by excluding workers who had substantial real estate or equity holdings. I faced equally difficult problems constructing h_t and s_t, but I will defer the details of their construction until later.

Good measurements of $\{x_t, d_t\}$ are absolutely critical to the success of the DP model since it is highly nonlinear in variables and there currently is no good theory of errors in variables for such models. Wherever possible, I have attempted to obtain independent measures of the variables to assess the magnitude of the measurement error. I have also constructed an array of associated "flag variables" to indicate the degree of confidence in each of the constructed state and control variables. By setting the appropriate flags, I can screen out questionable cases to obtain a core subsample for which confidence in the data is relatively high. To guard against the possibility that such screening could produce unpredictable sample selection biases, I have compared the distribution of each variable to its distribution in the full sample using all available observations. Because presentation of tabulations of the flag variables takes us too much into the "guts" of the computer programs that generate the state variables, I have decided against presenting them. Instead, I describe the nature of any special data or sample selection problems where appropriate.[4]

I can state the major conclusions of the data analysis as follows.

1. At the aggregate level, the data show workers making a smooth transition from work into retirement, gradually reducing consumption and labor supply but maintaining wealth levels intact. This is consistent with the behavior of a neoclassical, risk-averse consumer who attempts to smooth consumption and leisure streams and to provide bequests to his heirs. However, at the individual level, the data are anything but smooth: measured consumption shows erratic fluctuations, and labor supply has an abrupt discontinuity, with the typical worker staying at his full-time job up until retirement age (62–65), at which time he applies for Social Security, quits his job, and remains out of the labor force for the rest of his life.

2. Constructing consumption expenditures from the budget equation, $c_t = w_t - w_{t+1} + y_t$, is susceptible to frequent and often large measurement errors in wealth, possibly exacerbated by absence of good information on capital gains. The majority of the erratic variations in measured consumption appear to be attributable simply to response errors in wealth.

3. The distribution of real wealth changes is centered about zero, but with a large variance. On average, net worth is not very large, about four times annual income, and a substantial fraction of this wealth, 50–60 percent, is tied up in housing. These facts provide additional support for the view that the large swings in measured consumption are simply a result of response errors in wealth rather than erratic consumption/savings behavior. Although a simple test of the null hypothesis H_0: $c_t = y_t$ versus H_A: $c_t \neq y_t$ rejects at the 5 percent level (but not at the 1 percent level), the fact that the average change in wealth is \$−658 with a standard deviation of \$47,015 makes it very hard to distinguish between alternative theories of consumption/savings behavior. Because of the problems involved in accurately measuring wealth and therefore consumption, I have opted to start with a simpler DP model based on the hypothesis that $c_t = y_t$. In this model, workers choose labor force participation strategies to maximize the expected discounted value of the utility of income, abstracting from wealth and bequests.

4. Although respondent's total income is recorded only for even-numbered years, the existence of independent income measures in the SSER and SSMBR data sets allowed me to construct reliable income imputations in odd-numbered

years. Thus, if wealth changes are indeed an insignificant component of consumption, total imputed income will be a good measure of actual consumption.[5]

5. The distribution of total annual hours worked is highly bimodal, with most of its mass at either 0 or 2000. While some of this bimodality is likely an artifact of response error (with workers simply rounding their responses to forty hours/week, fifty weeks/year), it does indicate that the tripartite classification of labor force status e_t into 1 = full time, 2 = part time, or 3 = unemployed does not grossly misrepresent the data and that the measure is robust to fairly large variations in the hours cutoffs defining the three e_t states. Overall, the distributions provide little evidence to support the view that workers treat annual hours of work as a continuous decision variable.

6. A systematic response error problem known as the *seam problem* produces exaggerated estimates of labor state transitions across the survey dates, or seams, of the RHS. This leads to artificial cyclical variations in the transition probabilities for "across-seam" transitions as compared to "between-seam" transitions. The variation is apparently due to imperfect recall of labor force history in the earlier year of the two-year interview frame, leading to inconsistencies between recalled labor force status in the current interview and the labor force status reported in the last interview. One can ameliorate the seam problem by "skipping over the seams" and tracking transitions between the even-numbered years immediately preceding the odd-year survey dates in order to reduce the amount of recall on the part of respondents. This convinced me to formulate a DP model with a time period of two years rather than with a more fine-grained model with a one-year time period.

7. There are three possible measures of the "job search" control variable: SR, self-reported planned hours of work in the year following the survey; NE, actual hours worked in the year following the survey; and PC, actual hours worked in the year following the *subsequent* survey. The last measure corresponds to a "perfect control" model wherein an unemployed worker who decides to go back to work is successful with probability one. This analysis focuses on the other two measures, which correspond to "imperfect control" models where unemployed workers who decide to look for a full-time job have less than a 100 percent chance of being successful. Probably reflecting the fact that "talk is cheap," it appears that the first measure of s_t is a much more noisy measure of actual job search behavior than is the second measure. The data show that the second measure allows for a more intuitive and predictable relation between job search decisions and subsequent employment outcomes.

8. The four-way classification of health status h_t into 1 = good health, 2 = health limitation but not disabled, 3 = disabled, and 4 = dead seems to produce sensible results despite the inherently subjective nature of health status. Use of actual benefits paid from the SSMBR data was critical to the quality of h_t since self-reported measures of health significantly underestimate

the occurrence of health state 3 owing to systematic underreporting of Social Security disability receipts by respondents. The Social Security requirement of doctor examination for disability qualification seems to be a significant factor in identifying individuals with substantially greater health problems as indicated by their significantly higher ex post mortality. An unfortunate aspect of the disability classification is the fact that no workers become disabled after age 62. This is an artifact of Social Security rules that automatically convert disability payments into OASI payments after age 62.[6]

9. The SSMBR data allow me to identify when individuals actually apply for and receive OASI benefits. Thirty percent of eligible recipients apply for benefits as soon as they are able to receive them at the early retirement age 62, and another 30 percent apply for benefits at the normal retirement age 65. Overall, nearly all nondisabled workers apply for and receive Social Security retirement benefits between the ages of 62 and 65. The implied retirement hazard and frequency distributions computed using the SSMBR data and a definition of "retirement" as the age of first entitlement to OASDI differ significantly from the distributions computed by other researchers using the RHS data and other definitions of retirement. In order better to understand the phenomenon of early retirement and the pronounced bimodal distribution of retirement dates, I have incorporated a new control variable sr_t defined by

$$(2) \qquad sr_t = \begin{cases} 0 \text{ if worker has not applied for OASI,} \\ 1 \text{ if worker first applied for OASI} \\ \quad \text{before age 65 (early retirement),} \\ 2 \text{ if worker first applied for OASI} \\ \quad \text{after age 65 (normal retirement),} \end{cases}$$

and corresponding control variable ss_t defined by

$$(3) \quad ss_t = \begin{cases} 1 \text{ if worker applies for Social Security benefits,} \\ 0 \text{ if worker does not apply for Social Security benefits.}^{7} \end{cases}$$

Including sr_t and ss_t allows me to avoid *ad hoc* definitions of "retirement," separating the analysis of retirement behavior (i.e., collection of OASI) from labor supply behavior.

11.2 Estimation of Worker's Beliefs: Main Findings

The DP model represents workers' beliefs about uncertain future events by a Markov transition probability density $\pi(x_{t+1}|x_t, d_t)$. Under the assumption of homogeneity and rational expectations, one can "uncover" these beliefs from data on the realizations of $\{x_t, d_t\}$. Given the discretization of time and state variables proposed in section 11.1, π is a matrix with approximately 130 million elements. Even with my comparatively large data set of over thirty thousand observations on $\{x_t, d_t\}$, the standard nonparametric estimate of π is out of the question since nearly all cells of $\hat{\pi}$ would be estimated as identically zero even though workers might believe that the corresponding transitions occur with positive probability. Nonparametric approaches such as kernel and nearest-neighbor regressions also have problems since their estimates of π depend critically on arbitrary choices of kernel, window-width, and other smoothing parameters whose proper values I have little intuition about.[8] My approach is to find a parametric specification $\pi(x_{t+1}|x_t, d_t, \theta)$ that depends on a much lower-dimensional vector of unknown parameters θ in such a way that all relevant cells of π are assigned nonzero probabilities. It is also important to choose a specification that is parsimonious yet sufficiently flexible so that the estimated model is consistent with the data. Above all, it is crucial that the estimated beliefs are "sensible" if we expected to get "sensible" estimates of workers' preferences.

Direct parameterization of a 130 million element matrix seems out of the question, so a more clever approach must be employed. The strategy I have followed is to decompose π into a product of conditional and marginal densities and estimate each of the components separately. To see this more clearly, note that without loss of generality one can decompose a bivariate transition density f as follows:

$$
\begin{aligned}
(4) \qquad f(x_{t+1}, y_{t+1}|x_t, y_t) &= f_1(y_{t+1}|x_{t+1}, x_t, y_t)f_2(x_{t+1}|x_t, y_t) \\
&= f_3(x_{t+1}|y_{t+1}, x_t, y_t)f_4(y_{t+1}|x_t, y_t),
\end{aligned}
$$

where f_1, f_2, 3_3, and f_4 are defined from f in an obvious way. Although (4) shows that the ordering of the decomposition of f is irrelevant, it does make a difference when the functional form of f must be estimated from the data, especially where data measurement problems can lead to decompositions which exhibit "spurious causality." Having tried various decompositions of π, the one I found most plausible is given below

$$
\begin{aligned}
(5) \quad & \pi(y_{t+1}, e_{t+1}, sr_{t+1}, ms_{t+1}, h_{t+1}|y_t, e_t, sr_t, ms_t, h_t, a_t, s_t, ss_t) = \\
& \pi_y(y_{t+1}|e_{t+1}, ms_{t+1}, h_{t+1}, y_t, e_t, sr_t, ms_t, h_t, a_t, s_t, ss_t) \times \\
& \pi_e(e_{t+1}|ms_{t+1}, h_{t+1}, y_t, e_t, sr_t, ms_t, h_t, a_t, s_t, ss_t) \times \\
& \pi_{ms}(ms_{t+1}|h_{t+1}, y_t, e_t, sr_t, ms_t, h_t, a_t, s_t, ss_t) \times \\
& \pi_h(h_{t+1}|y_t, e_t, sr_t, ms_t, h_t, a_t, s_t, ss_t).
\end{aligned}
$$

Note that the decomposition (5) excludes the state and control variables c_t, w_t, aw_t from the original list presented in section 11.1. Consumption c_t and wealth w_t were excluded because of the measurement problems discussed in conclusion 3 of section 11.1. The Social Security average monthly wage aw_t (a complex average of the worker's historical earnings) was excluded since it turned out to be sufficiently collinear with current income y_t that I could reduce the dimensionality of the model by making y_t do double duty as a proxy for

aw_t. Finally, future age a_{t+1} and Social Security status sr_{t+1} were excluded since with probability one they obey trivial nonstochastic transition rules:

$$\begin{aligned} a_{t+1} &= a_t + 2, \\ sr_{t+1} &= sr_t \quad \text{if } sr_t \;\; 0, \\ sr_{t+1} &= 1 \quad \text{if } sr_t = 0,\ ss_t = 1,\ a_t < 65,\ a_t \geqslant 62, \\ sr_{t+1} &= 2 \quad \text{if } sr_t = 0,\ ss_t = 1,\ a_t \geqslant 65. \end{aligned}$$

The motivation for the decomposition of π given in (5) is that income y_t and employment status e_t are the most important state variables of the DP model, and therefore their evolution should be predicted as well as possible. If we view (5) as specifying π as a direct product of individual transition matrices, then π_y is the "innermost" component of the direct product, in the sense that income transitions are conditioned on the contemporaneously realized values of all the remaining state variables. From an empirical standpoint, including these contemporaneous values substantially improves the fit of the income regressions estimated in section 11.8.

The outermost component of the direct product, health status h_t, has additional structure resulting from the definition of health states $h_t = 3$ and $h_t = 4$. If I fix the values of the other variables (y_t, e_t, ms_t, a_t,d_t), then π_h represented by the 4 × 4 transition probability matrix:

$$(6) \qquad \pi_h = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{14} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{24} \\ 0 & 0 & \varphi_{33} & \varphi_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

According to (6), death is treated as an absorbing state. Note that disability is also treated as an absorbing state in the sense that, once a worker becomes disabled, he can only continue to stay disabled or die. This restriction was necessitated by data limitations. Although the Social Security SSMBR data set includes the variable "date of termination of disability benefits," there were only a handful of cases where actual termination was observed. Perhaps this indicates problems in Social Security record keeping, but it is more likely just an artifact of my Social Security–based definition of "disability." According to Social Security rules, disabled workers who receive SSDI benefits past age 62 are automatically reassigned OASI benefits after turning 62. Thus, there is no real incentive for Social Security to keep track of the date when the actual physical disability terminates once the worker is older than 62. One can try to partially rectify the problem the following way: reclassify workers who received disability benefits prior to age 62 and who are now older than 62 and reporting that they are in good health as being in state $h_t = 1$ rather than

$h_t = 3$. Unfortunately, this reclassification scheme has its own problems: although it allows transitions from disability to good health ($h_t = 3$ to $h_{t+1} = 1$), there is no way to record transitions from $h_t = 3$ to $h_{t+1} = 2$ since the RHS variables do not allow us to distinguish between the states "existence of a health problem that limits one's ability to work" and "disability."

The remaining sections of the paper discuss the construction of the state and control variables in more detail and present estimation results for each of the four components of the decomposition of π given in (5). Having conducted an extensive specification search to find an appropriate functional form for π, I can summarize the main findings below.

1. Age and income are relatively unimportant determinants of death rates after controlling for health, employment, and marital status. Death rates decrease slightly with income and actually decrease with age until age 67.[9] Not surprisingly, single workers are significantly more likely to die than married workers. However, even this variable has a small effect relative to health h_t and the labor supply/retirement decision (s_t, ss_t). Workers who are in poor health ($h_t = 2, 3$) are two to four times more likely to die than healthy workers. There is an equally strong association between the job search decision s_t and the probability of death, but the nature of the relation depends critically on the worker's health and retirement status. If the worker is retired or disabled ($h = 3$ *or* $ss \in \{1, 2\}$), any attempt to return to work on either a full- or a part-time basis is extremely hazardous, significantly increasing the risk of death. However, if the worker has not already retired and is in relatively good health ($h \in \{1, 2\}$ *and* $ss = 0$), the decision to quit work is associated with significantly higher death rates. Although this latter finding may represent spurious causality because of failure completely to control for all dimensions of health status, from the standpoint of a worker behaving according to the DP model the association is necessarily interpreted as cause and effect.

2. The probability of becoming disabled is a sharply decreasing function of age, a result that is an artifact of the definition of disability discussed above. It is clear that disability is an endogenous state variable (i.e., the outcome of an unmodeled underlying decision process), as evidenced by the fact that the probability of becoming disabled decreases well before age 62. The explanation is that the process involved in applying and qualifying for SSDI imposes significant costs on the worker, including doctor examination at the worker's expense. Naturally, the closer one is to the early retirement age of 62, the less incentive one has to incur the costs of applying for SSDI, especially when the probability of qualification is significantly less than one. Given the difficulty of constructing a sensible "objective" measure of disability, and given the fact that by law disabled workers have essentially no further labor supply/retirement decisions,[10] I have decided to treat disability/death as a combined absorbing state since the certification standards appear successfully to identify a group of workers who have serious health problems, as confirmed by *ex post* mortality rates which are twice as high as for nondisabled workers. Another

finding of interest is the fact that both single and higher-income workers are significantly less likely to become disabled.[11]

3. The probability of being in good health is a declining function of age and an increasing function of income. All other things being equal, marital status has no significant effect on the probability of being in good health. By far the most important determinant of future health is current health. Currently healthy workers are three times more likely to be in good health than currently unhealthy workers ($h_t = 2$). There is weak evidence that continuing to work on a full- or part-time basis is associated with a higher probability of being in good health. Conversely, the decision to quit working is associated with a deterioration in health. This result is corroborated by the fact that retired workers, $sr \in \{1, 2\}$, are significantly less likely to be in good health. As with my comments in point 1, the association might indicate spurious causality due to imperfections in the measure of health status: healthier workers continue working, while unhealthy workers quit and retire.

4. By far the most important variable predicting future marital status is current marital status: an older single worker has less than a 7 percent chance of finding a new mate over the two-year survey period. Older workers are more likely to lose their spouse, while higher-income workers are less likely to become single, at least up to an income of $30,000. There is weak evidence that, among single workers, the worse one's health, the more likely one is to remain single, although unhealthy married workers have a higher chance of remaining married. Economic decisions such as the labor search decision s_t or the retirement decision ss_t appear to have little or no effect on future marital status.

5. As one would expect, future employment status e_{t+1} is most strongly affected by the employment search decision s_t.[12] In addition, the worker's previous employment state e_t has a significant effect on probability that the search decision s_t is realized. Thus, currently employed workers who decide to continue working full time have a higher probability of remaining fully employed than part-time or unemployed workers. Interestingly, unemployed workers appear to have a significantly higher chance of being successful in gaining a full-time job than part-time employed workers. If a worker decides not to work, he is more likely to "realize" his decision if he is currently unemployed than if he had a full- or part-time job. Full-time workers are more likely to realize their quit decisions than part-time workers. Health status also has a very strong effect on employment status. Workers who become disabled are two and a half times more likely to be out of the labor force, and their chances of staying in a full-time job are less than one-third that of nondisabled workers. There are clear aging effects on the ability to continue working full time; for example, the probability that a 67-year-old worker will be successful in keeping or finding a full-time job is only one-third as high as that of an equivalent worker under 60. Income appears to be a statistically significant proxy for employability, with high-income workers being 60 percent more likely to keep or obtain a full-time job than

low-income workers. Somewhat surprisingly, changes in marital status have no significant effect on employment status. Less surprising is the fact that workers who are receiving OASI are less likely to be fully employed and more likely to be unemployed, all other things equal.

6. In order to match the long-tailed cross-sectional income distributions, the stochastic process for income was assumed to have a transition density with a conditionally heteroscedastic lognormal distribution. Income is strongly autocorrelated with an autoregressive coefficient of .95, and there is evidence of nonlinearity in this relation in the sense that higher powers of current income y_t enter the model with highly significant coefficients. The higher powers of y_t were needed primarily to enable the model to fit the complicated patterns of conditional heteroscedasticity that exist in the data. The estimated model has a variance of future income y_{t+1} that is an increasing function of current income, but the relation is far from proportional: a worker earning \$50,000 has a standard deviation in y_{t+1} of \$12,000, whereas a worker earning \$5,000 has a standard deviation in y_{t+1} of \$2,000. Health status has a significant effect on income prospects: healthy workers expect a 3 percent increase in real income, and disabled workers expect a 5 percent increase in income. However, currently healthy workers who become disabled expect a 20 percent drop in income. Changes in marital status have large and statistically significant effects on income. A worker who loses his wife expects a 25 percent drop in income, and a bachelor who has no prospects of remarriage expects his income to fall by about 20 percent. However, by far the most important determinant of future income y_{t+1} is the worker's employment status/search decision (e_t, s_t). Workers who keep working at their full-time jobs expect a 20 percent increase in income, while workers who exit from the labor force expect a 20–30 percent decrease in income.[13] The estimated income process successfully captures the main features of OASDI benefit rules, including the regressive nature of the payoffs, the extra benefits to a spouse, the early retirement penalty, and the effect of the "earnings test" for workers under 70.

7. It is possible that there exist unmeasured differences or *heterogeneity* among workers that create systematic differences in workers' beliefs but that are not captured in the list of state and control variables set forth in this paper. In order to assess the potential magnitude of this problem, I included several demographic variables in the estimation of workers' beliefs π, including a variable classifying the respondent as a "work lover" or "leisure lover" as well as his education, race, and the industry and occupation of his longest held job. Surprisingly, except for the finding that blacks expect significantly lower incomes than whites, none of these variables had a major effect on the estimated transition probability $\hat{\pi}$. Thus, the available evidence indicates that the list of state and control variables set forth in this paper provides a reasonably complete set of "sufficient statistics" for the states and decisions of my sample of workers. In particular, there is no compelling evidence that the failure to account for unmeasured heterogeneity leads to a gross misrepresentation of workers' beliefs.[14]

The remaining sections of the paper present the numerical evidence supporting the conclusions drawn in sections 11.1 and 11.2 and can be skipped or skimmed by readers who are willing to accept them at face value.

11.3 Analysis of Age, Marital Status, and Demographic Variables

A set of variables that we ought to be able to measure accurately are the identity of the respondent, his or her age, and basic demographic variables such as race, education, and the occupation/industry of the respondents' longest job. By and large this is true of the RHS data, although cross-checks of self-reported values with Census and Social Security records do indicate discrepancies. For example, out of an initial 1969 sample of 8,131 males, reported and recorded Census date of birth differed by more than one year in fifty-three cases, in some cases by more than ten years. In order to estimate the DP model, I need to track each of the 8,131 original male respondents over the ten-year survey period. RHS respondent identifiers allowed me to distinguish the original male respondent from his surviving spouse (or other household members), and, in conjunction with comprehensive death records compiled by Paul Taubman, I was able to determine whether the original respondent died, even if he was no longer responding to the survey. Table 11.1 provides a response summary that shows that the basic sample of original male respondents decreased from 8,131 in 1969 to 4,298 in 1979. There was significant attrition of the original 1969 male respondents over the survey. Table 11.1 shows that the attrition was due to the respondent's death in 2,327 cases and nonresponse in 1,506 cases. A discrepancy exists between the individual subrecord identifier in the SSER tapes and the respondent identifiers on the original RHS tapes: the former showed 8,091 original respondents in 1971 versus 7,054 in the RHS. The former figure could not possibly be right given that 433 respondents had died by the 1971 interview. Indeed, a second cross-check using the Census nonresponse file[15] agreed with the RHS identifiers. This provided a sobering reminder that one cannot necessarily trust the SSA's internal accounting data more than the RHS interview data.

Relatively minor discrepancies exist in the data on marital status. For example, nine individuals reported being married with spouse not present in

Table 11.1 **RHS Response Summary**

	71	73	75	77	79
Original '69 male respondent responds	7,054	6,239	5,541	4,811	4,298
No response, '69 respondent still alive	534	889	1,104	1,315	1,426
No response, '69 respondent dead	152	361	610	917	1,245
Surviving spouse responds, '69 respondent dead	244	488	722	908	1,075
Other relation responds, '69 respondent dead	37	58	54	60	7
Other relation responds, '69 respondent alive	110	96	100	120	80
Total	8,131	8,131	8,131	8,131	8,131

1969 but reported having never been married in 1971; two cases reported having a deceased spouse in 1969 and never having been married in 1971. Thirty-five cases classified themselves as being a surviving spouse in 1971 but listed themselves as having a "spouse in '69 but not in '71" instead of the correct response, "'69 spouse deceased, no '71 spouse." Using the corrected marital status data, I defined the marital state variable ms_t as follows:

$$ms_t = \begin{cases} 1 \text{ if respondent is married,} \\ 2 \text{ if respondent is widowed, separated, divorced, or never married.} \end{cases}$$

Table 11.2 presents the computed two-state Markov transition matrices for marital status (where M denotes cases that are missing owing to death or nonresponse). The transition matrices change in the expected way over time: the probability of becoming a widower over the two-year survey frame increases from 6 percent in 1969 to 9 percent in 1977. The probability of remarriage decreases over time from 7 percent in 1969 to 2 percent in 1977.

Table 11.3 presents the estimation results for π_{ms}, the marital status component of the decomposition of π given in (5). The elements of π_{ms} were estimated by maximum likelihood, using a linear-in-parameters, binomial logit specification of the probability that $ms_{t+1} = 2$.[16] The estimation results

Table 11.2 **Markov Transition Matrices for Marital Status**

Year of Transition	Cell Counts					Transition Probabilities	
	1	2	M	Total	%	1	2
1969–71:							
1	6,180	386	512	7,078	87	94	6
2	65	814	174	1,053	13	7	93
				8,131	100		
1971–73:							
1	5,434	405	406	6,245	84	93	7
2	66	976	158	1,200	16	7	93
				7,445	100		
1973–75:							
1	4,760	410	330	5,500	80	92	8
2	63	1,140	178	1,381	20	5	94
				6,881	100		
1975–77:							
1	4,141	391	312	4,844	75	91	9
2	48	1,310	215	1,573	25	4	96
				6,417	100		
1977–79:							
1	3,602	372	220	4,194	72	91	9
2	32	1,434	239	1,705	28	2	98
				5,899	100		

Table 11.3 **Estimates of Marital Status Transition Probability (dependent variable: $I\{ms_{t+1} = 2\}$)**

Variable	Estimate	t-statistic
$s_t = 1$	$-.05$	$-.3$
$s_t = 3$	$-.08$	$-.6$
$ms_t = 2, h_{t+1} = 1$	-1.87	-1.6
$ms_t = 2, h_{t+1} = 2$	-2.05	-1.8
$ms_t = 2, h_{t+1} = 3$	-2.68	-2.2
$ms_t = 1, h_{t+1} = 1$	4.06	3.6
$ms_t = 1, h_{t+1} = 2$	4.15	3.7
$ms_t = 1, h_{t+1} = 3$	4.67	4.2
a_t	$-.02$	-1.0
y_t	.15	8.5
$y_t \cdot y_t$	.002	-7.1
$ss_t \in \{1, 2\}$	$-.13$	$-.8$
$ss_t \in \{1, 2\}, ms_t = 2$	.02	.1
Log likelihood	$-2,347.8$	
Grad • direc	6 E-025	
Correctly predicted (%)	97	
Total observations	18,833	

in table 11.3 are based on a smaller subsample than table 11.2 (18,833 vs. 34,773 observations) as a result of conditioning on the availability of complete observations for the state and control variables entering π_{ms} and conditioning on a sample Boolean variable. The Boolean excludes respondents who are not the original 1969 male respondents and further excludes respondents who are farmers or farm owners, respondents with significant pension wealth, and respondents who made sufficiently erroneous or suspicious responses as determined from the flag variables described in section 11.1. Overall, the estimation results in table 11.3 support the conclusions drawn in point 4 of section 11.2.

11.4 Health Status

A key variable in the DP model is the worker's health status. This variable shifts the worker's mortality hazard and affects his ability to work and enjoy leisure. In order to construct the health status variable, I used mortality data from Paul Taubman's "death tape" and a battery of over seventy-five questions on health status in the RHS. It turned out, however, that two of the seventy-five RHS health variables were most relevant for classifying health status: HLIM, "Do you have any health condition, physical handicap, or disability that limits how well you get around?" and HWRK, "Does your health limit the kind or amount of work or housework you can do?" Originally, I used these variables, together with fifteen other health-related questions and the respondent's report as to whether he received SSDI benefits, to classify health status h_t into one of four states: 1 = respondent is in good health, 2 = respondent has a health problem that limits his ability to work or get around but is not severe enough for the worker to qualify for SSDI, 3 = respondent has a health problem severe enough for him to qualify for SSDI, and 4 = respondent is dead. My original construction of this variable yielded significantly lower estimates of the probability of being on SSDI than those of Bound (1986): 1.17 percent in 1969 versus Bound's estimate of 7.1 percent for men aged 55–64 in 1970. In addition, the data appeared to show an unexpected mass outbreak of poor health in 1975, with only 1,254 respondents classified as $h_t = 1$ and 3,958 classified as $h_t = 2$. By using the SSMBR OASDI payments data, I was able to directly verify whether a worker was classified as disabled by SSA by determining whether he qualified for SSDI payments. Furthermore, analysis of the health input variables revealed that the HWRK variable had 5,956 missing values in 1975 and that the remaining cases contained a disproportionate percentage of workers reporting a health limitation (1,476 out of 2,200). This turned out to be an artifact of a survey skip pattern introduced in 1975 that was different from skip patterns in other survey years: HWRK75 was asked only if respondent was in the labor force, whereas in other survey years the HWRK question was not conditioned on being in the labor force. I "fixed" the problem by using only the HLIM variable to classify workers into health state $h = 1$ or $h = 2$ and merging the disability data from the SSMBR to classify disabled workers $h = 3$.

Another problem arose from the fact that the RHS survey did not attempt to track workers who became institutionalized; it simply records them as missing. There is good reason to believe that the failure to track institutionalized workers induces a sample selection bias since single workers are less likely to have a family support network to rely on and are therefore more likely to become institutionalized and be lost from the sample. To correct this problem, I merged data from the Census nonresponse file, which records the reasons for nonresponse, including institutionalization. Analysis of health status of the institutionalized workers showed that among the sample of 113 institutionalized workers (36 percent of who were single in 1969 as compared to 13 percent for the sample as a whole), in only one case did the worker return to the RHS sample with improved health: the preponderant majority of institutionalized workers died within a few years after entering the institution. Based on this evidence, I decided to redefine health state 4 as an absorbing state for workers who are either dead, disabled, or institutionalized.

A final problem was more difficult to resolve. Although I have fairly complete data on the month and year that a worker died, in order to be included in the estimation of the health transition probability matrix, I must observe the worker's state and control vector (x_t, d_t) in the survey period immediately preceding his death. Unfortunately, there are many cases where the worker failed to respond to the survey for two or more survey periods preceding his death. Analysis of these cases shows that a disproportionate number consist of single men. One solution is to "remove" the intervening periods of missing data by treating the death as occuring just after the last survey to which the worker responded. Unfortunately, this approach has the effect of "accelerating" the deaths of a fairly large group of workers, distorting the estimates of age-death profiles. I decided, therefore, to leave the data as they were and simply acknowledge the possibility of sample selection bias that might lead to an underestimate of mortality rates for single workers.

Table 11.4 displays the transition probability matrices for my final definition of h_t. The data show a much more reasonable rate of disability receipt, 8.1 percent in 1969, which is much closer to Bound's estimate. The transition matrices generally appear to be quite reasonable, with workers in worse health states having significantly higher risk of death and disability. Mortality rates appear fairly stable over time and are in rough agreement with independent

Table 11.4 **Health Transition Probabilities**

	Cell Counts							Transition Probabilities			
Year	1	2	3	4	*M*	Total	%	1	2	3	4
1969–71:											
1	4,347	630	111	211	470	5,769	71	82	12	2	4
2	562	790	84	116	147	1,699	21	36	51	5	8
3	0	0	506	106	46	658	8	0	0	83	17
4	0	0	0	0	0	0	0	0	0	0	100
					8,126	100					
1971–73:											
1	3,629	730	76	181	296	4,912	70	79	16	1	4
2	449	688	51	126	107	1,421	20	34	52	4	10
3	0	0	533	113	56	702	10	0	0	83	17
4	0	0	0	0	0	0	0	0	0	0	100
					7,035	100					
1973–75:											
1	2,975	707	20	177	240	4,119	66	77	18	1	4
2	371	831	12	149	76	1,439	23	27	61	1	11
3	0	0	541	89	38	668	11	0	0	86	14
4	0	0	0	0	0	0	0	0	0	0	100
					6,226	100					
1975–77:											
1	2,495	510	0	164	217	3,386	61	79	16	0	5
2	422	877	2	171	87	1,559	28	29	59	1	11
3	0	0	454	92	43	589	11	0	0	83	17
4	0	0	0	0	0	0	0	0	0	0	100
					5,534	100					
1977–79:											
1	2,133	540	0	131	130	2,934	61	76	19	0	5
2	316	867	0	158	60	1,401	29	23	65	0	12
3	0	0	359	73	27	459	10	0	0	83	17
4	0	0	0	0	0	4	0	0	0	0	100
					4,798	100					

estimates calculated by Mott and Haurin (1985) using NLS data. Note that the transition probabilities in table 11.4 imply that disability is an absorbing state: once a worker becomes disabled, he either remains disabled, becomes institutionalized, or dies. This is simply a reflection of the data limitations discussed in section 11.2: the SSMBR data do not record the date of termination of disability. As a result, in each survey year there are approximately one hundred workers who report that they have no health problem that limited their ability to work or get around despite the fact that Social Security records indicate that they are disabled. Because the existing classification of disability confirms my a priori belief that disabled workers have significantly higher mortality rates, and, more important, because this classification matches the aggregate disability rates compiled by Bound, I decided not to reclassify these workers as $h_t = 1$.

Another apparent contradiction exists between Census/Social Security death records, the RHS death records, and the death records independently compiled by Paul Taubman. The RHS date of death differs from that in Taubman's data in thirty-six cases, which in turn differs from the Census and Social Security death date (from the SSMBR tape) in 302 cases. Case-by-case cross-checks resolved the discrepancies between Taubman's data and RHS, and cross-checks of Taubman's data with the Census data reveal that in 285 cases Taubman's data recorded the respondent as dead while Census and SSA had no record of death. Individual cross-checks reveal that Taubman's data are probably right in these cases. In fact, one can identify at least twenty-six cases of apparently fraudulent behavior involving a surviving spouse who continued to collect both her and her husband's OASI benefits even though the husband had been deceased for several years.[17] The final death data that I used to construct the health variable are Taubman's original data, edited in approximately sixty cases where case-by-case examinations revealed that either the RHS or the SSMBR death date was correct.

I conclude this section with tables 11.5–11.7, which present the estimates of the transition probabilities for health, disability, and death, respectively. Each of the transition probabilities was specified to have linear-in-parameters binomial logit functional forms. Products of the estimated probability functions can be multiplied out to compute the estimated health transition matrix, $\hat{\pi}_h$. The interpretation of the estimation results has been listed in points 1, 2, and 3 of section 11.2 and will not be repeated here. However, in order to get more intuition about how workers believe their health declines with age, I present figure 11.1, which shows $\text{pr}\{h_{t+1}=1|a_t\}$, $\text{pr}\{h_{t+1}=3|a_t\}$, and $\text{pr}\{h_{t+1}=4|a_t\}$, respectively.

The age-health profile graph in figure 11.1 shows the probability of remaining in good health as a function of age for four configurations of the remaining state variables. All four curves show that health declines with age; however, changes in the other variables have a stronger effect on health than age alone. The top two curves (marked with circles and squares) represent the health expectations of workers who are already in good health and who retire

Table 11.5 **Estimates of Health Transition Probability (dependent variable $I\{h_{t+1} = 1\}$)**

Variable	Estimate	t-statistic
$h_t = 1, s_t = 1$	−2.28	−4.7
$h_t = 1, s_t = 2$	−2.15	−4.3
$h_t = 1, s_t = 3$	−1.84	−3.7
$h_t = 2, s_t = 1$	−.13	−.2
$h_t = 2, s_t = 2$	−.21	−.4
$h_t = 2, s_t = 3$	.20	.4
a_t	.01	1.6
y_t	−.05	−7.2
$y_t \bullet y_t$	.001	5.2
$ms_t = 2$	.02	.3
$ss_t \in \{1, 2\}$	.17	2.9
Log likelihood	−8,470.7	
Grad • direc	2 E-028	
Correctly predicted (%)	79	
Total observations	17,536	

Table 11.6 **Estimates of Disability Hazard Function (dependent variable: $I\{h_{t+1} = 3\}$)**

Variable	Estimate	t-statistic
Constant	−27.83	−12.7
$h_t = 1, s_t = 1$	.68	2.2
$h_t = 1, s_t = 3$	−.39	−1.0
$h_t = 2, s_t = 1$	−.65	−2.0
$h_t = 2, s_t = 2$	−.73	−1.7
$h_t = 2, s_t = 3$	−.95	−2.6
a_t	.51	14.9
y_t	.41	3.0
$a_t \bullet y_t$	.00	−3.0
$ms_t = 2$	.73	2.5
Log likelihood	−1,048.4	
Grad • direc	7 E-27	
Correctly predicted (%)	99	
Total observations	17,763	

at ages 62 and 65, respectively (the latter worker also has 10 percent higher income). The bottom curve represents the health expectations of a worker who is in poor health, $h_t = 2$, and who retires at age 62. The remaining curve, marked with x's, shows the health expectations of a healthy worker who continued to work until age 70, at which time he fell ill ($h_t = 2$), quit his job, and began collecting OASI. The combination of all these events at age 70 produced a dramatic downturn in the worker's health outlook.

Table 11.7 **Estimates of Mortality Hazard Function (dependent variable: $I\{h_{t+1} = 4\}$)**

Variable	Estimate	t-statistic
$h_t = 1, s_t = 1$	3.02	14.0
$h_t = 1, s_t = 2$	1.67	3.6
$h_t = 2, s_t = 1$	2.47	7.0
$b_t = 2, s_t = 2$	2.35	2.3
$h_t = 2, s_t = 3$	.20	.5
$h_t = 3, s_t = 1$	−1.83	−7.4
$h_t = 3, s_t = 2$	−.93	−3.4
$h_t = 3, s_t = 3$	−.33	−2.6
y_t	.01	1.5
$a_t \in [0, 60)$	2.15	15.7
$a_t \in [60, 62)$	2.35	20.1
$a_t \in [62, 65)$	2.71	26.5
$a_t \in [65, 68)$	3.03	28.1
$a_t \in [68, 71)$	2.92	24.3
$a_t \geq 71$	2.79	15.2
$ms_t = 2$	−.31	−4.3
$h_t = 1, s_t = 1, ss_t \in \{1, 2\}$	−3.15	−15.0
$h_t = 1, s_t = 2, ss_t \in \{1, 2\}$	−1.82	−3.9
$h_t = 1, s_t = 3, ss_t \in \{1, 2\}$	−.27	−2.5
$h_t = 2, s_t = 1, ss_t \in \{1, 2\}$	−3.40	−9.6
$h_t = 2, s_t = 2, ss_t \in \{1, 2\}$	−3.14	−3.1
$h_t = 2, s_t = 3, ss_t \in \{1, 2\}$	−.84	−2.3
Log likelihood	−4,713.9	
Grad • direc	2 E-027	
Correctly predicted (%)	94	
Number of observations	24,233	

Figure 11.1 also shows the probability of becoming disabled as a function of age. In this case, age effects dominate, reflecting sharp declines in workers' incentives to incur the costs of applying for disability benefits as they approach the early retirement age, 62. The topmost curve corresponds to a low-income married worker who is currently in poor health ($h_t = 2$), while the lowest curve corresponds to a high-income single worker who is in good health.

Finally, figure 11.1 plots the estimated death hazard function. As discussed in section 11.2, it was difficult to identify the independent effect of age on death rates. Both linear and quadratic specifications of age effects produced ultimately falling death hazards, a result I found implausible. Using age dummies, I discovered the explanation: the age dummies reveal that workers' death rates decrease until age 67, after which they begin rising with age. However, because the RHS surveyed men between 58 and 63 in 1969, the oldest possible age reached during the survey is 73. This implies that there are relatively few observations beyond age 67, so that both the linear and the quadratic specifications attempted to fit the downward-sloping part of the death

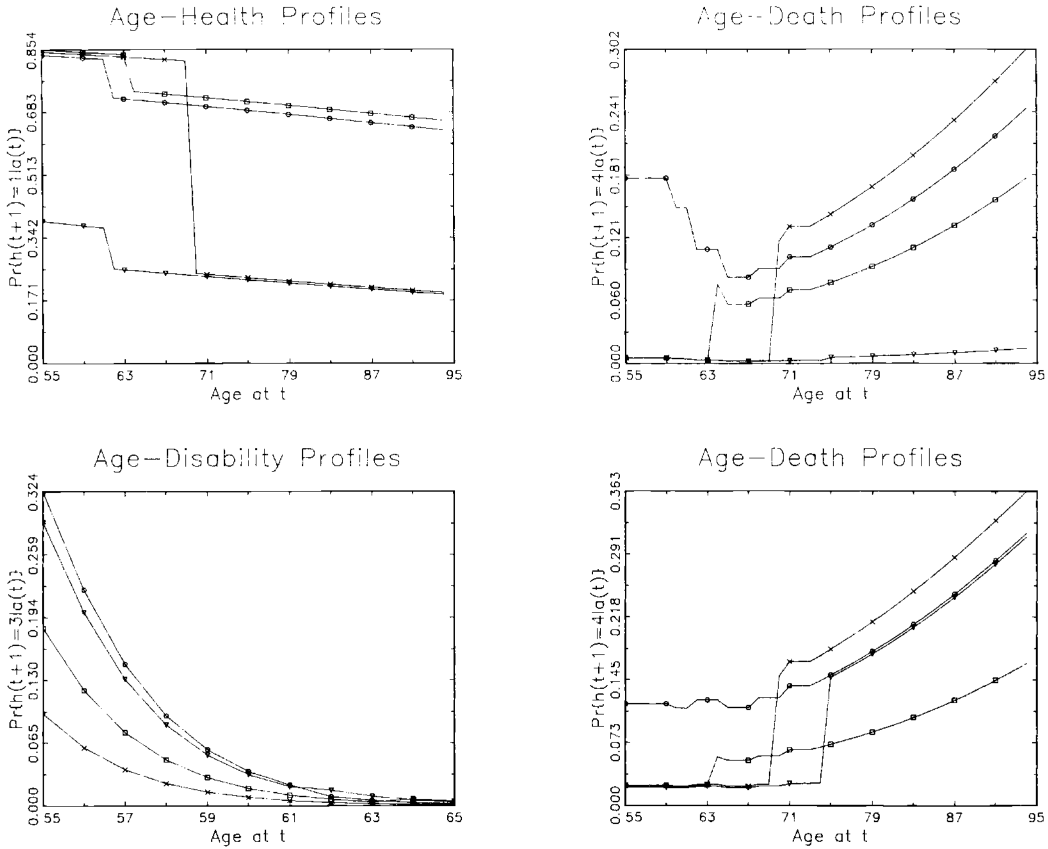

Fig. 11.1 Predicted health as a function of age

hazard function from age 58 to age 67, ignoring the upturn that occurred afterward owing to a lack of observations. Unfortunately, while the age dummies allow the model to fit the data well, it implies that the risk of death is constant after the worker reaches his early 70s. Aggregate mortality statistics show (unconditional on health and employment status) that death rates increase with age, which implies that a model containing only age dummies from 58 to 73 will ultimately underpredict death rates. To correct this, I added a pure age trend to the model in order to match the aggregate mortality statistics from age 74 to age 95.[18] Figure 11.1 (which incorporates the age trend after age 73) displays mortality expectations for four different workers. The V-shaped curve marked with circles corresponds to a single, low-income disabled worker. While his death rate is much higher than average, it shows significant improvement until age 67, after which it begins to worsen steadily. The bottom curve, marked with triangles, shows the mortality expectations of a high-income "workaholic" who is in good health and who continued working full time until his health deteriorated to $h_t = 2$ at age 75, after which he started working part time. In spite of his health problems, the workaholic never retired, in the sense of collecting OASI. The remaining two curves (marked with squares and x's, respectively) show the death rates of two average-income workers who retire at 65 and 70, respectively. The latter worker retired at 70 owing to the fact that his wife died and his health deteriorated from $h = 1$ to $h = 2$; this explains the dramatic increase in his death rate.

There are two features of figure 11.1 that seem implausible: the sharp V-shaped death hazard for the disabled worker and the significantly lower death rates for the high-income "workaholic" in comparison to the two average-income workers who retired at 65 and 70. Looking back to the estimation results in table 11.7, it appears that these predictions result from the fact that Social Security recipients ($sr_t \in \{1, 2\}$) have significantly higher death rates. As discussed in conclusion 1 of section 11.2, this is probably due to the fact that h_t does not capture all dimensions of health status. Workers who are in worse health are probably more likely to retire than healthy workers. However, from the standpoint of the DP model, the relation is necessarily treated as cause and effect: it implies that collection of OASI can be hazardous to your health. To avoid this problem, I reestimated the model without the sr_t interactions. While there was a significant drop in the log-likelihood (from $-4{,}714$ to $-5{,}045$), the graph in the lower-right-hand corner of figure 11.1 shows that the resulting model seems to produce more reasonable predictions. In particular, the age effects now show a slightly increasing rather than decreasing hazard rate, and the gross disparities between the workaholic (who never collected OASI) and his average-income colleagues have disappeared. Based on these results, I have decided to exclude the sr_t interactions in the specification of the mortality hazard, even though they clearly improve the fit of the model.

11.5 Employment Status

Accurate classification of employment status e_t is the key to the entire undertaking: employment status is the most important variable affecting income and utility levels in the DP model and is a crucial input into the income imputation routines that construct biennial income. They are also key inputs for the construction of biennial consumption expenditures in section 11.7. The RHS data set allowed me to construct three independent measures of labor force status: self-classification of employment status (SE), instantaneous employment status (IE), and historical employment status (E). Each of the measures assumes three values, 1 = full time, 2 = part time, and 3 = not employed. The SE variable was directly recorded in a trichotomous format from the survey question "Do you consider yourself partly retired, completely retired, or not retired at all?" The IE measure was determined from the survey question "How many hours per week do you usually work on your current job?" Using this response, I defined IE = 1 if the worker worked more than twenty-five hours per week, IE = 2 if the worker worked between five and twenty-five hours per week, and IE = 3 if the worker was not currently employed or worked less than five hours per week. The historical employment status measure E is an annual measure based on the total number of hours worked in the preceding year. I defined E = 1 if the respondent worked more than 1,300 hours in the past year, E = 2 if the respondent worked between 200 and 1,300 hours, and E = 3 otherwise. Because the worker might have had multiple jobs in the two years preceding the RHS interview, computation of total hours worked required direct reconstruction of the underlying continuous-time labor force histories from a battery of more than 130 questions in the "Work Experience" section of the RHS survey. Previous studies have used the IE and SE measures of employment status, probably because they were among the easiest variables to pull off the RHS tapes. Constructing retrospective labor force histories is a considerably more complicated undertaking owing to the existence of complicated skip patterns in the survey questionnaire and the need carefully to account for the beginning and ending dates of jobs when there are multiple transitions within the interview frame. To my knowledge, this is the first study to construct complete labor force histories using the RHS data.

Table 11.8 presents aggregate employment distributions using each of the definitions of employment status. Although there are significant differences between the measures, all three confirm conclusion 1 of section 11.1 that the aggregate data show workers making a smooth transition from work into retirement. The main differences are that SE appears to overestimate substantially the occurrence of part-time work relative to the E and IE measures, and E appears to overestimate part-time work relative to the IE measure. The latter effect is to be expected from the nature of the definition of E: a worker who worked at a full-time job until mid-year and then retired would be classi-

Table 11.8 Cross-sectional Distributions of Measures of Employment Status

	Historical Employment Status (%)										
	E68	E69	E70	E71	E72	E73	E74	E75	E76	E77	E78
1	71	72	61	54	40	34	23	19	14	13	10
2	9	9	12	13	15	15	16	15	14	14	14
3	20	19	27	33	45	51	61	66	72	73	76
N	8,117	7,379	7,379	6,837	6,837	6,392	6,392	5,871	5,871	5,415	5,415

	Instantaneous Employment Status (%)					
	IE69	IE71	IE73	IE75	IE77	IE79
1	74	61	40	24	16	12
2	4	5	7	9	10	10
3	22	34	53	67	74	78
N	8,117	7,434	6,897	6,392	5,871	5,415

	Self-reported Employment Search Decision (%)					
	SR69	SR71	SR73	SR75	SR77	SR79
1	72	56	36	22	13	10
2	5	12	13	15	13	12
3	23	32	51	63	74	78
N	7,894	7,434	6,897	6,392	5,871	5,415

	Self-Employment Status (%)					
	SE69	SE71	SE73	SE75	SE77	SE79
1	77	60	36	21	12	9
2	8	12	16	18	19	18
3	15	28	48	61	69	73
N	8,070	7,431	6,881	6,387	5,861	5,407

fied as being in state 2 by the E measure and in state 3 by the IE measure.

The SE variable seems like the poorest candidate for use as a measure of employment status owing to the ambiguity of the term "retired." Some people may interpret being "retired" as quitting their career job and will report being fully or partly retired even though they are working full time at a new job. Other people may interpret "retired" as meaning "are you working now?" and will report that they are not retired if they had quit their main career job but are currently working at a new full-time job. Still others may report being partly retired even though they are not working because they like to think that they have the virility to return to work at some unspecified future date. The latter problem seems to be reflected in table 11.8, which shows that the SE measure substantially overestimates the incidence of part-time work, sometimes as much as 200 percent in comparison to the IE measure. I decided not to use the SE measure because of the problems of ambiguity and subjective interpretation and also, for reasons I elaborate below, because SE is an "instantaneous" measure that does not correspond well to the time intervals of the DP model.

The instantaneous employment status IE variable completely avoids the subjective definition of the concept "retired." Like SE, IE has a high response rate and is easy to pull from the tapes. However, it too has certain drawbacks from the standpoint of estimating the DP model. Since I am using a relatively coarse two-year time interval (for computational reasons discussed in sec. 11.1), the instantaneous IE measure would not provide a good measure of the worker's actual state over the whole time period. In principle, workers may have changed jobs many times in the two-year time interval or may have retired only recently, so there may be only a weak association between IE and the respondent's actual labor force status over the last two years.

From the standpoint of the discrete-time DP model, the most appropriate measure of labor force status is the historical employment status measure, E. The main drawbacks of this measure are that (1) it requires the worker to recall his employment history (which may be especially difficult in the cases where the worker had multiple job transitions) and (2), since E is a flow measure, it may overestimate the occurrence of part-time work by misclassifying full-time workers who retire in mid-year. Table 11.9 sheds some light on the last problem by summarizing the distribution of employment histories (using the E measure of e_t) over the eleven years of the RHS survey. To keep the table manageable, the 11^4 possible employment sequences have been "collapsed;" for example, the sequence (1, 1, 1, 2, 2, 3, 3, 3, 3, M, M) (where M represents missing data) is classified as a "1–2–3" sequence.[19]

The first thing to notice is that, in contrast to the aggregate employment statistics, the individual employment sequences are far from smooth: only 18 percent of the sample is observed to phase out of work gradually in a "1–2–3" employment sequence. If I reclassify all "1–2–3" sequences with only one intervening year in state 2 as actually being a misclassified "1–3" sequence, then only 3 percent of the sample is observed to follow a smooth employment transition; a plurality of the sample, 33 percent, are observed to follow the discontinuous "1–3" sequence. Another 28 percent of the sample have complex "nonmonotonic" employment histories, with periods of unemployment followed by subsequent reemployment. Of course, many of the "1" and "1–2" sequences may actually be right-censored sections of an ultimate "1–2–3" sequence; however, since these sequences account for only 14 and 4 percent of the sample, respectively, accounting for censoring will not change the basic picture.

For comparison, table 11.9 presents the distribution of employment sequences for the IE and SE measures of e_t and also for an annual measure of

Table 11.9 Distribution of Employment Sequences

Measure of Employment State	Sequence	Cases N	Cases %
E	1 . . .	1,174	14
	2 . . .	91	1
	3 . . .	1,033	13
	1–3 . . .	1,488 (2,700)	18 (33)
	2–3 . . .	255	3
	1–2 . . .	306	4
	1–2–3 . . .	1,450 (238)	18 (3)
	Others	2,334	29
	Total	8,131	100
IE	1 . . .	1,321	16
	2 . . .	28	1
	3 . . .	1,337	16
	1–3 . . .	3,269	40
	2–3 . . .	112	1
	1–2 . . .	276	4
	1–2–3 . . .	308	4
	Others	1,480	18
	Total	8,131	100
SE	1 . . .	1,239	15
	2 . . .	131	2
	3 . . .	897	11
	1–3 . . .	2,642	33
	2–3 . . .	298	4
	1–2 . . .	601	7
	1–2–3 . . .	748	9
	Others	1,575	19
	Total	8,131	100
NLS[b]	1 . . .	585	23
	2 . . .	13	1
	3 . . .	187	7
	1–3 . . .	1,052	42
	2–3 . . .	29	1
	1–2 . . .	90	4
	1–2–3 . . .	89	4
	Others	452	18
	Total	2,497	100

Note: Numbers in parentheses obtained by reclassifying all "1–2–3" sequences with only one intervening year in state $e_t = 2$ as a "1–3" sequence. See sec. 11.4 for further explanation.

[a]An annual measure of employment status similar to E. This measure was constructed by Berkovec and Stern (1989), who wrote more than 2,000 lines of Fortran code to accurately follow NLS skip patterns to reconstruct the employment histories.

e_t similar to the E measure but computed from the NLS data by Berkovec and Stern (1989). Notice that, in all the tables, only 3–4 percent of all workers are observed to follow a "1–2–3" sequence. The NLS data show a somewhat higher fraction of workers following a "1" sequence, but this is to be expected given that the NLS sample follows a younger group of men, who were initially aged 45–59 in the first year of the survey, 1966.[20] Based on the comparison of the employment measures presented in table 11.9, I decided to reclassify all "1–2–3" employment sequences with only one intervening year in state 2 as a "1–3" sequence by reassigning the state $e_t = 2$ as either $e_t = 1$ or $e_t = 3$, depending on whether hours worked in that year are greater than 1,000.

Overall, table 11.9 casts doubt on the notion that most workers gradually phase out of their full-time jobs through a spell of "partial retirement," a view promoted by Gustman and Steinmeier (1984) and suggested from casual interpretation of the macro data in table 11.8. Even if I counted all "1–2" and "1" sequences as forming part of an eventual "1–2–3" sequence, the number of "smooth" employment transitions would be at most 23 percent. In reality, most of the "1" sequences will form part of an eventual "1–3" sequence, and a large fraction of the "3" sequences are actually left-truncated "1–3" sequences. If I count all these sequences as "1–3" sequences, I obtain an estimate that approximately 75 percent of all retirement sequences involve discontinuous transitions from a full-time job into unemployment. Table 11.9 also shows that a significant fraction of the sample, over 18 percent, follow "nonmonotonic" sequences involving some form of "unretirement," that is, a return to full employment from a state of unemployment or partial employment. Table 11.10 provides more detail on the structure of the nonmontonic employment sequences for the E, IE, and SE measures of employment status. The structure of these transitions is extremely complex, as can be seen from table 11.10. The most common nonmonotonic sequences are "3–1–3," "1–3–2," "1–3–1," "1–3–2–3," and "1–3–1–3." Even though a majority of workers follow the "1–3" sequence, the traditional approach to modeling retirement behavior as an *ex ante* choice of a fixed retirement date after which the worker ceases to work is incapable of explaining the labor force history of at least 20 percent of the sample.

The discussion above suggests the possibility that the discretization of the labor force status variable into just three states could seriously misrepresent the labor force participation decision. Other researchers (e.g., MaCurdy 1983) have suggested that the labor force participation decision can be modeled as a continuous choice variable, say, as choice of annual hours of work. There are strong practical reasons for maintaining this viewpoint: an interior solution allows one to derive stochastic Euler orthogonality conditions that permit estimation of identified parameters by the method of moments (Hansen 1982). Figure 11.2, which displays the distribution of annual hours of work over the period 1968–78, provides convincing evidence against this view. The distribution

Table 11.10 **Distribution of Nonmonotonic Employment Sequences**

	IE			E			SE		
Sequence	*N*	% of 1,480	% of 8,131	*N*	% of 2,334	% of 8,131	*N*	% of 1,575	% of 8,131
1313	99	6.69	1.22	36	1.54	.44	39	2.48	.48
1312	28	1.89	.34	20	.86	.25	15	.95	.18
1213	47	3.18	.58	75	3.21	.92	46	2.92	.57
1212	33	2.23	.41	83	3.56	1.02	71	4.51	.87
1231	10	.68	.12	21	.90	.26	24	1.52	.30
1232	28	1.89	.34	101	4.33	1.24	85	5.40	1.05
1321	30	2.03	.37	30	1.29	.37	18	1.14	.22
1323	100	6.76	1.23	115	4.93	1.41	158	10.03	1.94
2131	1	.07	.01	2	.09	.02	0	.00	.00
2132	1	.07	.01	11	.47	.14	7	.44	.09
3131	12	.81	.15	6	.26	.07	0	.00	.00
3132	8	.54	.10	9	.39	.11	5	.32	.06
3231	2	.14	.02	3	.13	.04	0	.00	.00
3232	6	.41	.07	10	.43	.12	8	.51	.10
3121	5	.34	.06	5	.21	.06	2	.13	.02
3123	17	1.15	.21	90	3.86	1.11	8	.51	.10
3213	2	.14	.02	15	.64	.18	3	.19	.04
3212	0	.00	.00	7	.30	.09	3	.19	.04
2121	3	.20	.04	10	.43	.12	2	.13	.02
2123	15	1.01	.18	77	3.30	.95	26	1.65	.32
2321	2	.14	.02	3	.13	.04	1	.06	.01
2323	7	.47	.09	34	1.46	.42	34	2.16	.42
2312	1	.07	.01	0	.00	.00	1	.06	.01
2313	3	.20	.04	3	.13	.04	1	.06	.01
131	122	8.24	1.50	35	1.50	.43	78	4.95	.96
121	105	7.09	1.29	122	5.23	1.50	109	6.92	1.34
132	201	13.58	2.47	137	5.87	1.68	204	12.95	2.51
321	12	.81	.15	12	.51	.15	7	.44	.09
312	19	1.28	.23	25	1.07	.31	8	.51	.10
323	57	3.85	.70	80	3.43	.98	68	4.32	.84
313	176	11.89	2.16	102	4.37	1.25	35	2.22	.43
213	34	2.30	.42	75	3.21	.92	46	2.92	.57
231	2	.14	.02	5	.21	.06	4	.25	.05
232	5	.34	.06	22	.94	.27	24	1.52	.30
212	5	.34	.06	20	.86	.25	25	1.59	.31
31	77	5.20	.95	51	2.19	.63	60	3.81	.74
32	40	2.70	.49	45	1.93	.55	43	2.73	.53
21	9	.61	.11	30	1.29	.37	24	1.52	.30
Others	156	10.54	1.92	807	34.58	9.92	283	17.97	3.48
Total	1,480	100.00	18.20	2,334	100.00	28.70	1,575	100.00	19.37

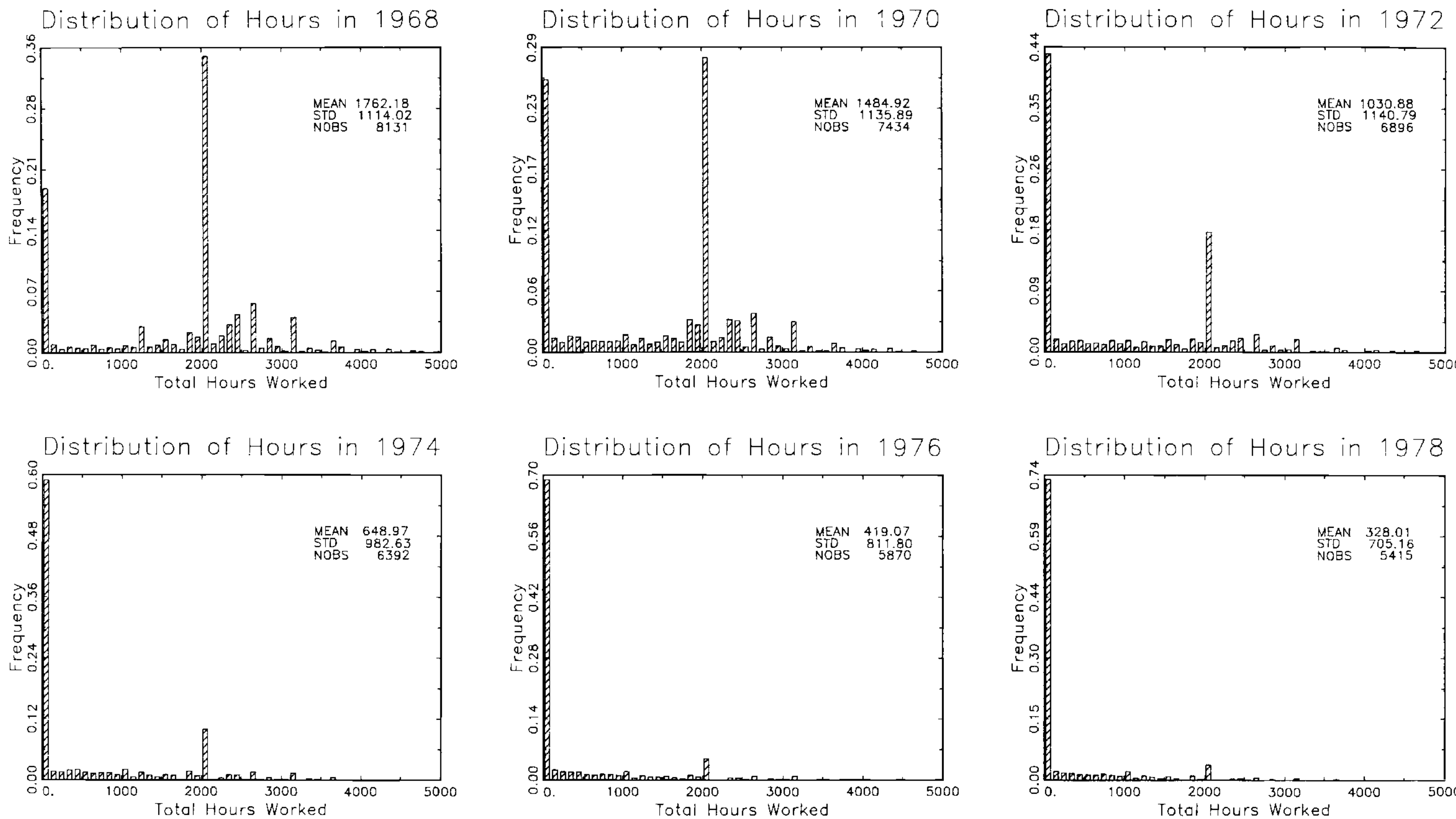

Fig. 11.2 Distributions of annual hours of work, 1968–78

has almost all its mass at two spikes, one at zero and the other at 2,000. The distributions are almost excessively concentrated at the two spikes, suggesting a systematic tendency of respondents to round their reponses (e.g., forty hours/week, fifty weeks/year). Nevertheless, I believe that the distributions provide solid evidence against the notion that annual hours of work is best modeled as a continuous choice variable that satisfies an interior first-order condition. This notion is also supported by the work of Gustman and Steinmeier (1983, 1984), who present convincing evidence of widespread minimum hours restrictions and significant wage cuts associated with transitions from full-time work to part-time. These constraints, combined with the Social Security "earnings test," are probably the key factors that lead the majority of workers to follow a "bang-bang" work/no work decision rule. Figure 11.2 also shows that the definition of the E variable is robust to fairly large changes in the cutoffs defining the three employment states: there is a small amount of probability mass uniformly distributed between zero and 2,000 hours of work, so that changes in the cutoff in this range will not significantly alter the distribution of e_t.

I conclude this section by presenting Markov transition probability matrices for the IE and SE measures of e_t in table 11.11 and the E measure in table 11.12. These are "uncontrolled" transition probabilities because I have not conditioned on a measure of the respondent's search decision, s_t. Nevertheless, the resulting transition matrices are quite illuminating. The matrices show a clear pattern of age effects: for example, table 11.11 shows that the probability of reemployment (i.e., a transition from $e_t = 3$ to $e_{t+1} = 1$) is 16 percent in 1969 but falls to 2 percent by 1977 and that the probability of remaining fully employed declines from 75 percent in 1969 to 55 percent in 1977. Interestingly, the probability of retiring from a full- or part-time job peaks at approximately 40 percent between 1971 and 1973, declining to 30 percent in 1977.

A strange pattern appears in the historical employment state transition matrices in table 11.12. Notice how the transition matrices appear to cycle in two-year intervals: for example, the (1, 1) elements appear significantly higher in even-numbered years, while the (3, 3) elements appear significantly higher in the odd-numbered years. For a long time, I was convinced that these regular fluctuations had to be an artifact of my FORTRAN code for processing the observations. I labored for many weeks to make sure that my program accurately followed the complicated skip patterns in the survey questionnaire but had no success in eliminating the strange fluctuations in the transition matrices. Only recently have I become aware of work by Daniel Hill (1988) that has convinced me that the fluctuations are not artifacts of my computer programs but rather symptoms of a systematic response error problem known as the *seam problem*. The seam problem arises from the way the RHS collects data on retrospective labor force history in successive two-year survey frames. Each of the odd-numbered survey years represents a seam, and the survey

Table 11.11 Transition Matrices for IE and SE Measures of Employment Status

IE69 to IE71 (N = 8,117)			SE69 to SE71 (N = 8,070)		
76	4	20	73	11	16
30	37	33	19	42	39
16	3	81	6	6	88
IE71 to IE73 (N = 7,434)			SE71 to SE73 (N = 7,431)		
59	6	35	55	15	30
19	39	42	12	45	43
6	3	91	4	7	89
IE73 to IE75 (N = 6,897)			SE73 to SE75 (N = 6,881)		
52	9	39	50	17	33
15	46	39	9	54	37
3	4	93	2	7	91
IE75 to IE77 (N = 6,392)			SE75 to SE77 (N = 6,387)		
53	11	35	45	23	32
15	51	34	7	54	39
3	4	93	2	6	92
IE77 to IE79 (N = 5,871)			SE77 to SE79 (N = 5,861)		
55	15	30	52	22	26
18	48	34	10	56	34
2	4	94	2	6	92

questionnaire required the respondent to recall his labor force history in the two-year survey frame prior to the interview. It appears that, while respondents offer an internally consistent view of the preceding two years, their view of history changes between survey dates in a way that generates inconsistent labor force transitions across seams. For example, to compute the across-seam transition probability matrix from 68 to 69, I needed data from two different surveys: the 1969 survey gave me retrospective data on labor force states in 68, and the 1971 survey gave me retrospective data on labor force states in 69. On the other hand, the between-seam transition probability matrix from 69 to 70 was computed entirely from retrospective data obtained at the 1971 interview. The pattern of fluctuations in the transition matrices indicates that men in state $e_t = 1$ are more likely to remain in state 1 for across-seam transitions than for between-seam transitions, whereas men in state $e_t = 3$ are less likely to remain in state 3 for across-seam transitions than for between-seam transitions.

I have recomputed table 11.12 using the flag variables to eliminate observations that showed any evidence of internally inconsistent responses.

Table 11.12 Markov Transition Matrices for Historical Employment Status

E68 to E69 (seam: N = 8,117)			E73 to E74		
91	3	6	65	14	21
30	43	27	9	61	30
23	8	69	0	2	98
E69 to E70			E74 to E75 (seam: N = 6,392)		
82	7	11	70	10	20
18	58	24	14	58	28
2	1	97	2	5	93
E70 to E71 (seam: N = 7,379)			E75 to E76		
82	5	13	63	16	22
29	51	20	9	65	27
5	7	88	1	2	97
E71 to E72			E76 to E77 (seam: N = 5,871)		
71	10	19	76	9	15
11	59	30	11	64	25
1	2	97	1	5	94
E72 to E73 (seam: N = 6,837)			E77 to E78		
74	8	18	66	22	12
21	57	22	11	67	22
3	7	90	0	3	97

While the sample sizes were significantly reduced, the seam problem persisted. Although an analysis of the perceptual/psychological factors underlying the seam problem is beyond the scope of this paper, it appears that, by using between-seam transitions based on data from a single survey frame, one is much more likely to obtain a consistent set of transition probabilities. Indeed, looking at the between-seam transition matrices in table 11.12 one can see that they change in a sensible way over time, with no suspicious patterns indicative of further inconsistencies. In particular, while the transition matrices do not closely match the IE or SE transition matrices (the latter two are two-year transition matrices, while the E transition matrix is for one-year intervals), the matrices follow the same general pattern as the IE and SE transition matrices, namely, a probability of reemployment and continued employment that gradually declines over time. I conclude that the seam problem is sufficiently severe to make it inadvisable to build a DP model based on annual data even though such a model is superior from a theoretical viewpoint since it has "finer grain" and thus suffers less from problems of time aggregation. Instead, I will focus on constructing a model of biennial transitions, using consistent data on employment transitions between seams rather than across seams.

11.6 Job Search Decision

The DP model requires a control variable s_t that represents the respondent's labor force search/participation decision. In a discrete-time model, the agent is in labor force state e_t at time t, and, conditional on e_t and his search decision s_t, he makes a transition to a new labor force state e_{t+1} at time $t + 1$. Thus, the DP model gives an employed worker the option of quitting ($e_t = 1$ or $e_t = 2$, and $s_t = 3$) and an unemployed worker the option of returning to work ($e_t = 3$, and $s_t = 1$ *or* $s_t = 2$). Unfortunately, while it is convenient to trichotomize s_t into three values (1 = search for full-time job, 2 = search for part-time job, and 3 = quit the labor force), the "true" search decision is essentially a latent variable: a complicated, possibly multidimensional variable encompassing the variety and intensity of each of the worker's possible search activities over the period. In the RHS, there are three possible variables from which to construct a measure of s_t:

SR: self-reported planned hours of work in the year following the survey,
NE: actual hours worked in the year following the survey,
PC: actual hours worked in the year following the *subsequent* survey.

The latter measure corresponds to a perfect control DP model where workers' search decisions are successful with probability one. The PC measure may seem implausible given the well-known labor market problems of older workers, yet on the other hand it necessarily suffers much less from measurement error. In this paper, however, I focus on the other two measures of s_t.

Using the SR and NE measures, I constructed a trichotomous estimate of s_t using the same cutoffs that I used to construct the e_t variable described in section 11.5. Table 11.8 summarizes aggregate distribution of the self-reported measure of s_t. This measure follows very much the same trends as the E, SE, and IE measures of e_t: a gradual phase-out from full employment into unemployment. The NE measure of s_t is recorded in the odd-year columns of the E distribution at the top of table 11.8. At least on the aggregate level, the two measures appear to track each other fairly closely.

To get a better handle on the issue of which measure of s_t better approximates the underlying latent employment search decision, I computed the controlled transition probability matrices that predict the probability of e_{t+1} conditional on e_t *and* s_t. Table 11.13 presents the controlled transition matrices using the E measure for e_t and the SR measure for s_t. These matrices show a very weak relation between employment search decisions and *ex post* realized employment states. If control were perfect, the transition matrix should have ones in the column corresponding to the value s_t assumes. However, in table 11.13 we see that under the SR measure control is highly imperfect. For example, a full-time worker who reported an intention to quit working in 1969 still has a 25 percent chance of remaining at work in 1971. A worker who had a full-time job in 1968 and who reported an intention to start working part time

Table 11.13 **Controlled Transition Probabilities, SR Measure of Job Search Variable, s_t**

E68 to E70 Given SR69 = 1 (N = 5,707)			E72 to E74 Given SR73 = 1 (N = 2,508)		
81	7	12	60	14	26
58	20	22	40	34	26
50	17	33	27	28	45
E68 to E70 Given SR69 = 2 (N = 394)			E72 to E74 Given SR73 = 2 (N = 859)		
33	13	54	12	20	67
24	51	25	14	55	31
3	20	77	4	35	61
E68 to E70 Given SR69 = 3 (N = 1,793)			E72 to E74 Given SR73 = 3 (N = 3,470)		
26	12	62	3	6	91
14	20	66	5	20	75
15	5	80	2	4	94
E70 to E72 Given SR71 = 1 (N = 4,150)			E74 to E76 Given SR75 = 1 (N = 1,372)		
71	10	19	58	14	28
54	25	21	31	36	33
29	29	42	13	16	71
E70 to E72 Given SR71 = 2 (N = 884)			E74 to E76 Given SR75 = 2 (N = 981)		
4	20	76	7	20	73
14	54	32	11	55	34
5	34	61	6	22	72
E70 to E72 Given SR71 = 3 (N = 2,345)			E74 to E76 Given SR75 = 3 (N = 4,039)		
5	6	89	1	11	88
8	17	75	3	17	80
2	4	94	2	4	94

in 1969 has only a 20 percent chance of actually realizing his intentions by 1970. An unemployed worker in 1974 who reports the intention to return to work full time in 1975 has only a 13 percent chance of actually being employed in 1976. Thus, the SR measure of s_t leads to a DP where control is too *imperfect,* in the sense that there is an implausibly low correspondence between employment search decisions and subsequent labor market outcomes.

Table 11.14 presents controlled transition probabilities for the E measure of e_t and the NE measure of s_t. Comparing tables 11.13 and 11.14, we can see that, while the NE measure of s_t does reflect imperfect control, the relation between s_t and e_{t+1} is much stronger than for the SR measure of s_t. For example, consider the probability that a worker who intends to quit his full-time job is successful (i.e., the transition from $e_t = 1$ to $e_{t+1} = 3$ given

Table 11.14 Controlled Markov Transition Probabilities, NE Measure of Employment Search Decision, s_t

E68 to E70 Given E69 = 1 (N = 5,348)			E72 to E74 Given E73 = 1 (N = 2,183)		
83	6	11	67	12	21
76	15	9	48	36	16
78	11	11	61	21	18
E68 to E70 Given E69 = 2 (N = 554)			E72 to E74 Given E73 = 2 (N = 817)		
14	58	28	7	54	39
23	57	20	11	61	28
12	60	28	7	69	24
E68 to E70 Given E69 = 3 (N = 1,477)			E72 to E74 Given E73 = 3 (N = 3,335)		
3	2	95	1	1	98
1	4	95	1	10	89
1	1	98	0	2	98
E70 to E72 Given E71 = 1 (N = 3,767)			E74 to E76 Given E75 = 1 (N = 1,167)		
72	9	19	66	13	21
55	25	20	38	39	23
67	20	13	60	13	27
E70 to E72 Given E71 = 2 (N = 645)			E74 to E76 (Given E75 = 2 (N = 790)		
10	54	36	4	49	47
14	56	30	10	67	23
6	71	23	7	70	23
E70 to E72 Given E71 = 3 (N = 2,376)			E74 to E76 Given E75 = 3 (N = 3,914)		
1	3	96	1	5	94
4	13	83	2	9	89
0	1	99	0	2	98

$s_t = 3$). In 1968, the NE measure gives a 95 percent chance that the decision will be realized, compared to only 62 percent for the SR measure of s_t. In the case of an unemployed worker who intends to return to work, the data for 1974 show that, according to the NE measure of s_t, the worker will have a 60 percent chance of success, compared to only a 13 percent chance for the SR measure of s_t. It is perhaps not surprising that the NE measure of s_t should have a strong correspondence with e_{t+1} since s_t is simply a lagged value of e_{t+1} and the $\{e_t\}$ process is highly serially correlated. However, it is somewhat surprising that the SR measure of s_t has such a weak correspondence with subsequent employment outcomes. This may be an indication of the fact that "talk is cheap": it is one thing to say that you intend to remain employed or return to work but quite another thing actually to go out and do it. A model using the

NE measure is a compromise between the implausible perfect control model implied by the PC measure of s_t and the perhaps equally implausible imperfect control model implied by the SR measure.

Tables 11.15 and 11.16 present the maximum likelihood estimates of the controlled transition probabilities using the E measure of e_t and the NE measure of s_t. The estimates correspond to the component π_e in the decomposition of π given in (5). The probabilities were estimated using a

Table 11.15 Estimates of Employment Status Transition Probability (dependent variable: $I\{e_{t+1} = 1\}$)

Variable	Estimate	t-statistic
$e_t = 1, s_t = 1$	3.64	9.15
$e_t = 2, s_t = 1$	2.50	6.06
$e_t = 3, s_t = 1$	2.96	7.01
$e_t = 1, s_t = 2$	−.92	−2.06
$e_t = 2, s_t = 2$	−.22	−.51
$e_t = 3, s_t = 2$	−1.12	−2.38
$e_t = 1, s_t = 3$	1.66	2.18
$e_t = 2, s_t = 3$	.85	1.15
$e_t = 3, s_t = 3$	1.42	1.98
$h_t = 1, h_{t+1} = 1$	.04	.45
$h_t = 1, h_{t+1} = 2$	−.09	−.79
$h_t = 1, h_{t+1} = 3$	−.86	−2.60
$h_t = 2, h_{t+1} = 1$	−.01	−.07
$h_t = 2, h_{t+1} = 3$	−.50	−1.32
$h_t = 3, h_{t+1} = 3$	−.82	−2.43
$a_t \in [0, 60)$	.46	3.17
$a_t \in [60, 62)$	.37	3.01
$a_t \in [62, 65)$	.16	1.56
$a_t \in [65, 68)$	−.05	−.50
$y_t \in [0, 4)$	−.52	−2.60
$y_t \in [4, 7)$	−.55	−3.23
$y_t \in [7, 10)$	−.45	−2.69
$y_t \in [10, 13)$	−.43	−2.50
$y_t \in [13, 21)$	−.34	−1.92
$y_t \in [21, 31)$	−.15	−.73
$ms_t = 2, ms_{t+1} = 2$	−.10	−.29
$ms_t = 1, ms_{t+1} = 2$	−.17	−.44
$ms_t = 1, ms_{t+1} = 1$	−.21	−.64
$ss_t \in \{1, 2\}, s_t = 1$	−1.50	−15.96
$ss_t \in \{1, 2\}, s_t = 2$	−.76	−3.47
$ss_t \in \{1, 2\}, s_t = 3$	−2.09	−3.42
Log likelihood	−9,154.9	
Grad • direc	7 E-028	
correctly predicted (%)	82	
Number of observations	18,778	

Table 11.16 **Estimates of Employment Status Transition Probability (dependent variable: $I\{e_{t+1} = 3\}$)**

Variable	Estimate	t-statistic
$e_t = 1, s_t = 1$	$-.52$	-1.24
$e_t = 2, s_t = 1$	-1.19	-2.67
$e_t = 3, s_t = 1$	$-.81$	-1.77
$e_t = 1, s_t = 2$	-1.52	-2.97
$e_t = 2, s_t = 2$	-1.87	-3.66
$e_t = 3, s_t = 2$	-1.89	-3.65
$e_t = 1, s_t = 3$	2.75	4.09
$e_t = 2, s_t = 3$	1.70	2.58
$e_t = 3, s_t = 3$	3.49	5.35
$h_t = 1, h_{t+1} = 1$	$-.44$	-5.15
$h_t = 1, h_{t+1} = 2$	$-.02$	$-.25$
$h_t = 1, h_{t+1} = 3$	.73	2.48
$h_t = 2, h_{t+1} = 1$	$-.38$	-3.03
$h_t = 2, h_{t+1} = 3$	.77	2.20
$h_t = 3, h_{t+1} = 3$	.48	2.53
$a_t \in [0, 60)$	$-.12$	$-.77$
$a_t \in [60, 62)$	.24	1.98
$a_t \in [62, 65)$	.41	4.30
$a_t \in [65, 68)$	.08	.86
$y_t \in [0, 4)$	.27	1.28
$y_t \in [4, 7)$	.20	1.00
$y_t \in [7, 10)$	.32	1.62
$y_t \in [10, 13)$	.44	2.11
$y_t \in [13, 21)$	.53	2.51
$y_t \in [21, 31)$	.32	1.28
$ms_t = 2, ms_{t+1} = 2$	$-.27$	$-.81$
$ms_t = 1, ms_{t+1} = 2$	$-.42$	-1.10
$ms_t = 1, ms_{t+1} = 1$	$-.41$	-1.28
$ss_t \in \{1, 2\}, s_t = 1$	1.03	6.80
$ss_t \in \{1, 2\}, s_t = 2$	1.08	3.45
$ss_t \in \{1, 2\}, s_t = 3$	.46	.87
Log likelihood	$-9,154.9$	
Grad • direc	7 E-028	
Correctly predicted (%)	82	
Total observations	18,778	

linear-in-parameters specification of a trinomial logit model of the probability that e_{t+1} assumes the three values $\{1, 2, 3\}$. Table 11.15 presents the parameter estimates corresponding to the event $I\{e_{t+1} = 1\}$ (full-time work), while table 11.16 presents the parameter estimates corresponding to the event $I\{e_{t+1} = 3\}$ (unemployment).[21] The interpretation of the estimation results has already been summarized in conclusion 5 of section 11.2 and will not be repeated here.

11.7 Income, Wealth, and Consumption

Next to employment status, the most important state variables of the DP model are income y_t and wealth w_t. The RHS has detailed information on assets and debts in each of the odd-numbered survey years, 1969–79, as well as detailed information on the components of income in the preceding even-numbered years, 1968–78. Although consumption c_t is treated as an observable control variable, in reality it is essentially a time aggregration of thousands of individual unobserved buy/no buy decisions over the two-year period. My strategy was to use the budget equation $w_{t+1} = w_t + y_t - c_t$ to infer consumption expenditures from measurements of w_{t+1}, w_t, and y_t. There are two obstacles to this approach: the RHS has no data on capital gains income, and the RHS records income only in even-numbered years. Thus, capital gains and income in odd-numbered years must be imputed. A key to accurate income imputations is the use of the retrospective labor force histories to construct the e_t state variable.

I initially tried to impute the missing income values by regressing income in even-numbered years on variables available in both even- and odd-numbered years. Among the variables available in both even and odd years were the SSER earnings records (up until 1974) and the SSMBR OASDI benefit data (from 1969 to 1978). Despite the inclusion of these variables and retrospective data on total hours worked in odd-numbered years, the fits of the income regressions were not very impressive, with R^2 values of 60 percent at best. Using the estimated regressions to fill in the missing income values produced intuitively unreasonable results, generating wide swings in income that occasionally turned negative or exceeded reasonable values.

An approach that turned out to work much better was a simple *ad hoc* procedure I call "full information interpolation." One can divide income into four sources: (1) wage income, (2) OASDI income, (3) unemployment insurance, and (4) other income. Since I have OASDI income in all years, that variable does not need to be imputed.[22] In addition, since other income is predominantly asset and pension income, which is largely independent of labor force participation, I obtained an estimate for category 4 by simply averaging observed other income in adjacent even-numbered years. The problem thus reduced to computing wage income and unemployment compensation. Using the retrospective employment histories, I obtained an estimate of total hours worked in each year. Dividing hours worked into observed wage income, I obtained a wage rate that I used to compute total wage income in odd-numbered years.[23] If there was evidence that the worker had become involuntarily unemployed during the period, I imputed unemployment compensation as well. The resulting interpolation estimates appeared much more reasonable than the regression-based imputations. In particular, there were far fewer wild swings in income, very few excessively large values, and no negative income values. Figure 11.3 plots the imputed and reported income distributions for the six-year period 1973–78. There is evidently little difference between the imputed and the reported income distributions; both have the characteristic lognormal shape. There is a noticeable leftward shift in the distribution over time as more and more workers withdraw from the labor force. This shift is not as pronounced as it might be because of the replacement of wage income by OASDI and pension receipts. If I were to plot wage distributions only, the leftward shift would be much more pronounced.

The existence of the seam problem in the employment data discussed in section 11.4 led me to suspect the possibility that these inconsistencies might have contaminated the imputed income data. To see whether there was any evidence of this, I plotted the distributions of income changes in figure 11.4. These distributions show no evidence of the seam problem, perhaps because wage income became an increasingly less important source of income over the survey and because the SSER earnings records and the SSMBR OASDI benefit data allowed me to get relatively accurate measurements of the main components of income for the majority of the sample. In any event, I conclude that my income imputations appear to be fairly reliable measures of actual income.

Having said this is not to deny the existence of systematic response errors in reported wage and OASDI benefits. For example, section 11.3 discussed the widespread underreporting of Social Security disability benefits. To assess how accurately respondents reported their income, I used the SSER and SSMBR data sets to compare reported and actual earnings and OASDI benefits. Because of the Social Security maximum earnings limitation, OASDI recipients had a clear incentive to deny or underreport their wage earnings since the survey was conducted for SSA. On the other hand, OASDI benefits themselves do not enter into the "earnings test," so there is no obvious incentive to underreport these receipts. Figure 11.5 presents the distribution of the percentage difference between reported wages and SSER earnings in 1970 and the distribution of percentage response error in total OASDI benefit in 1974.[24] The figure shows no obvious evidence of systematic underreporting, although each contains spikes at −100 percent indicating a nonnegligible fraction of respondents falsely reporting that they had no wage or OASDI income. On the basis of these comparisons, I set flags indicating the degree of accuracy of the respondent's reports of his wage and OASDI benefits. I then used these flags in the construction of a sample boolean to screen out questionable respondents.

I used the Hurd wealth data (see n. 2 above) to compute respondents' net worth. Net worth consists of financial and real assets less total indebtedness, but excludes pensions, life insurance, and annuities (the latter two are fairly uncommon in the RHS anyway).[25] Wealth data are extremely hard to cross-check because major components of wealth, such as the market value of the respondent's house, are often subjective guesses. Figure 11.6 plots the distribution of wealth for the six survey years. Notice that there is a significant

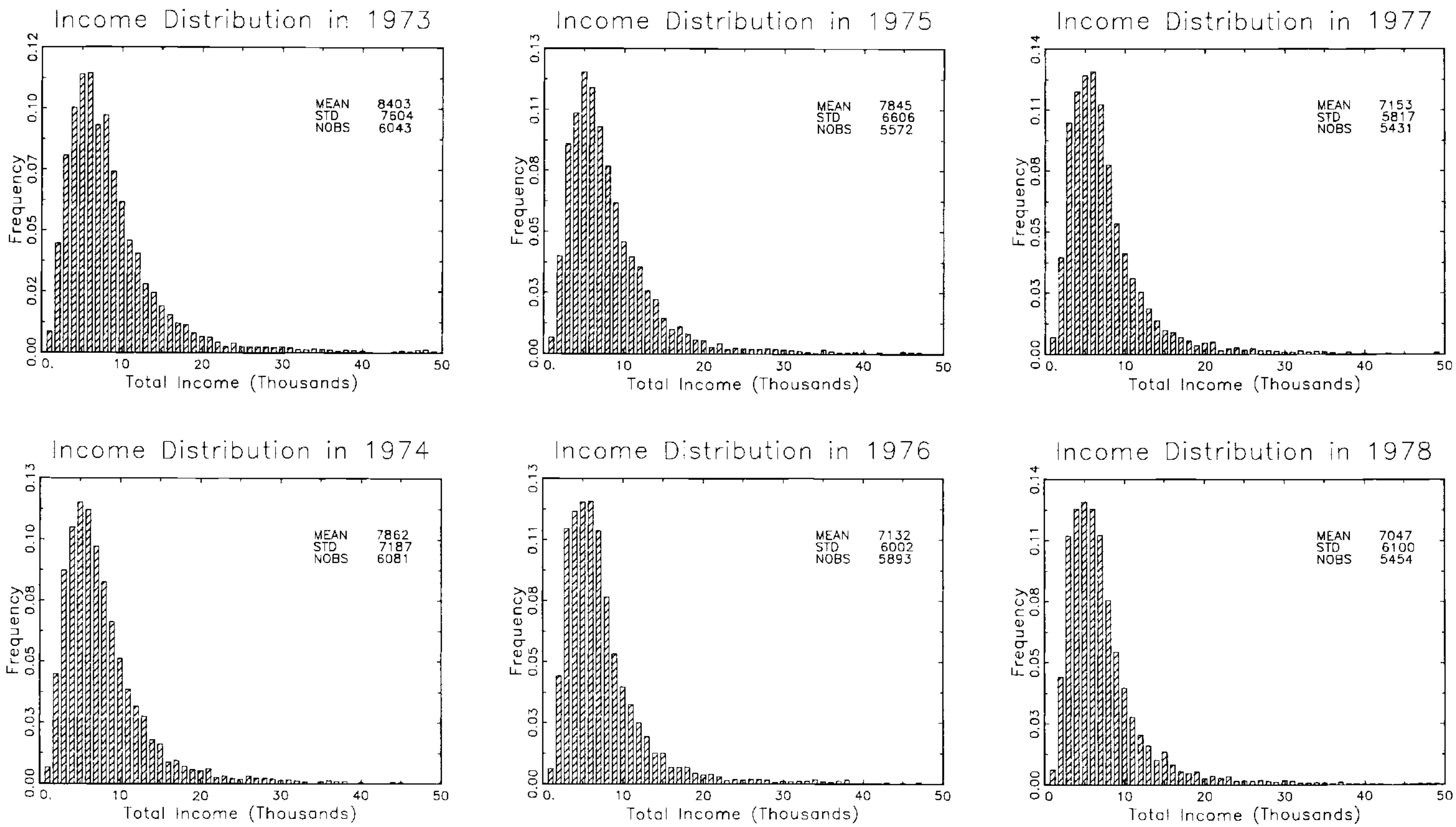

Fig. 11.3 Distributions of actual and imputed income, 1973–78

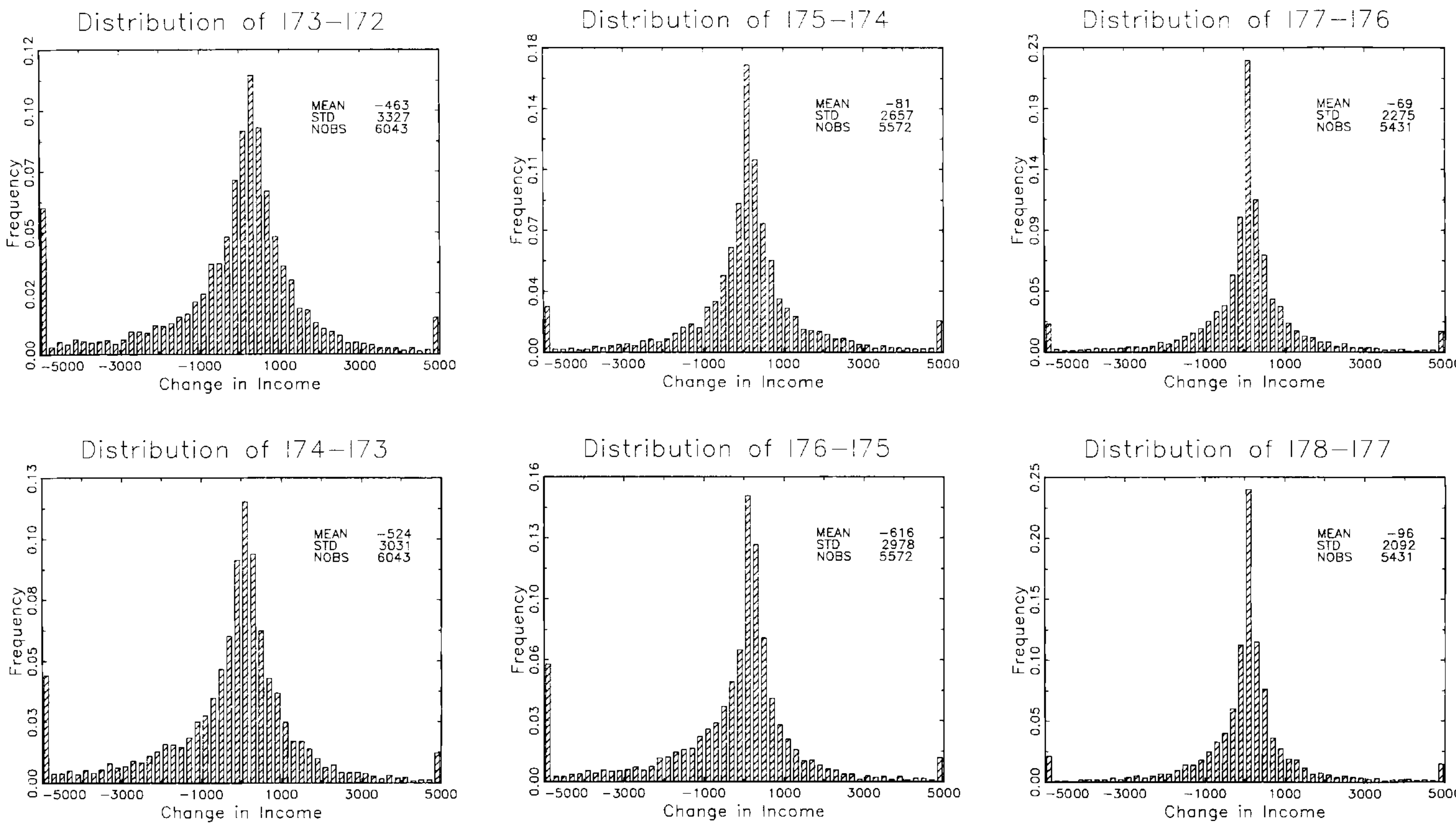

Fig. 11.4 Distributions of changes in income, 1973–78

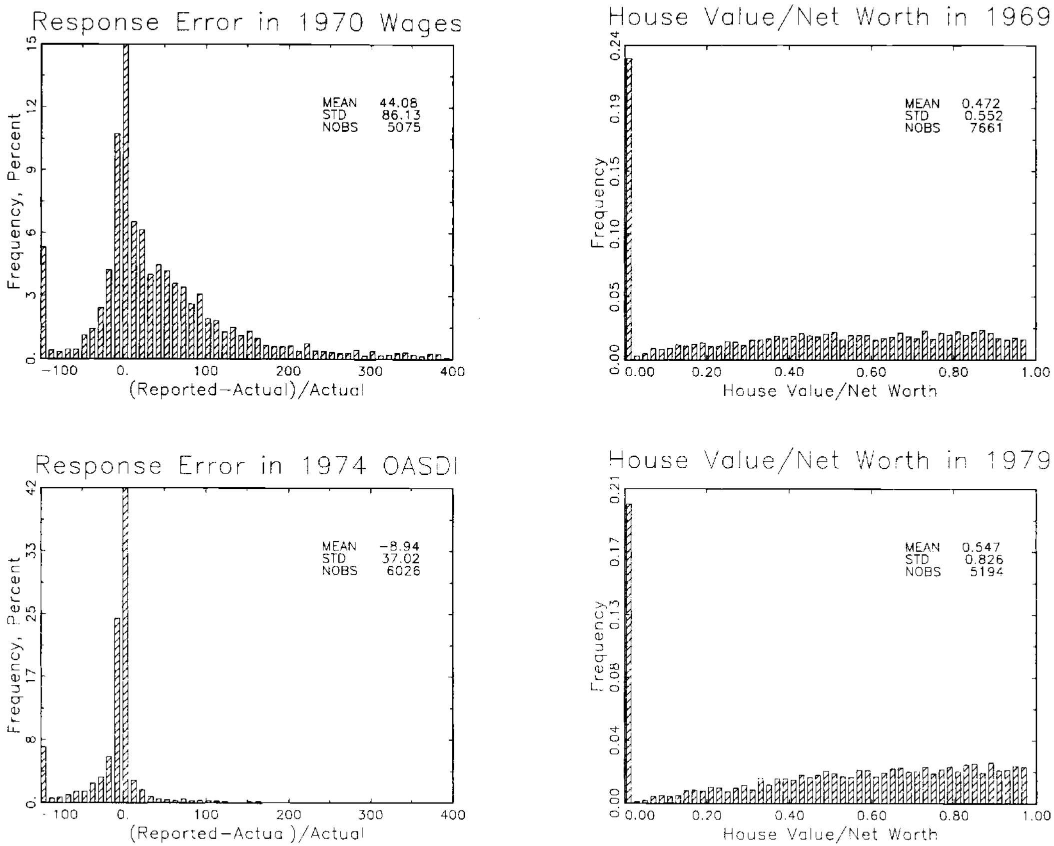

Fig. 11.5 Distribution of response error and ratio of net housing equity to net worth

fraction of respondents, about 10 percent, who report that they have essentially no tangible wealth. Mean wealth levels are about $28,000 1968 dollars, equal to approximately four years of income. These distributions provide little evidence that respondents consume their wealth as they age. Figure 11.5 plots the distribution of housing value to net worth in 1969 and 1979. It shows that a large fraction of workers' wealth is tied up in housing: homeowners have an average 56 percent of their wealth tied up in housing in 1969, increasing to 65 percent in 1979. The failure of wealth to decrease over time may be partly due to the appreciation of housing in the inflationary 1970s.

Using Hurd's wealth data and my imputed income series, I constructed an imputed biennial consumption series using the budget identity $c_t = w_t - w_{t+1} + y_t$. The resulting consumption distributions are plotted in figure 11.7. Overall, the distribution of consumption looks very similar to the distribution of income plotted in figure 11.8; both income and consumption show a noticeable tendency to shift leftward over time. This fact is not an accident since figure 11.9 shows that the distribution of wealth changes is centered about zero, suggesting that to a first approximation, $c_t = y_t$. Indeed, the mean wealth change (averaged over all periods and workers) is $\$-658$, with a *standard deviation* of $47,015. Given that average wealth is $28,000, it is difficult not to conclude that most of the variation is due to measurement error. The large standard deviation suggests that it would be difficult to reject the hypothesis that $c_t = y_t$. However, a simple hypothesis test of H_0: $c_t = y_t$ versus H_A: $c_t \neq y_t$ yields a χ^2 statistic of 6.2 with a marginal significance level of just over 1 percent: a rejection that is perhaps not surprising given that I have 31,348 observations on wealth changes.[26]

Whether the large variance in wealth reflects explainable differences in behavior or simple measurement error is an open question, but my initial investigations suggest the dominance of the latter. Like the employment data, aggregate consumption appears fairly smooth, slowly declining over time in apparent accord with the standard life-cycle hypothesis. However, at the individual level, measured consumption is anything but smooth, making violent, unpredictable swings over time. Overall, a total of 1,984 respondents have negative measured consumption in at least one of the five biennial survey periods, and in successive periods more than half the sample is recorded as having either a consumption increase of more than 200 percent or a consumption decrease of more than 50 percent. These large swings in consumption fly in the face of intuition and personal observation of the consumption behavior of the elderly, suggesting that most of the swings are due to measurement errors in wealth.

One possible reason for negative consumption is the failure to account for capital gains. Given the subjectivity of respondents' assessment of housing values and the fact that a majority of workers continue to live in the same house rather than "size down," it seemed reasonable to attribute all changes in net housing wealth to capital gains (provided the respondent did not move).

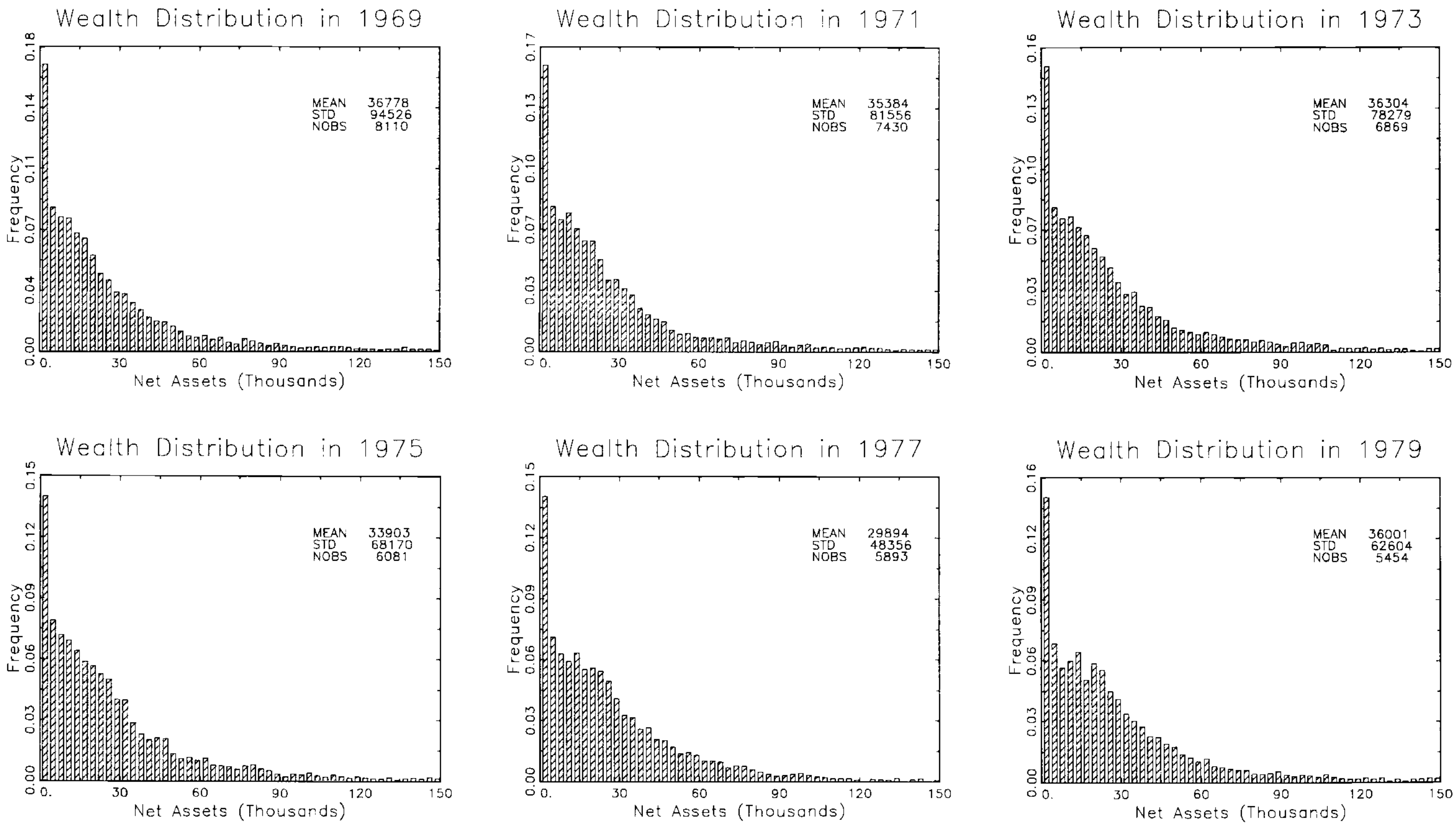

Fig. 11.6 Distribution of net worth, 1969–79

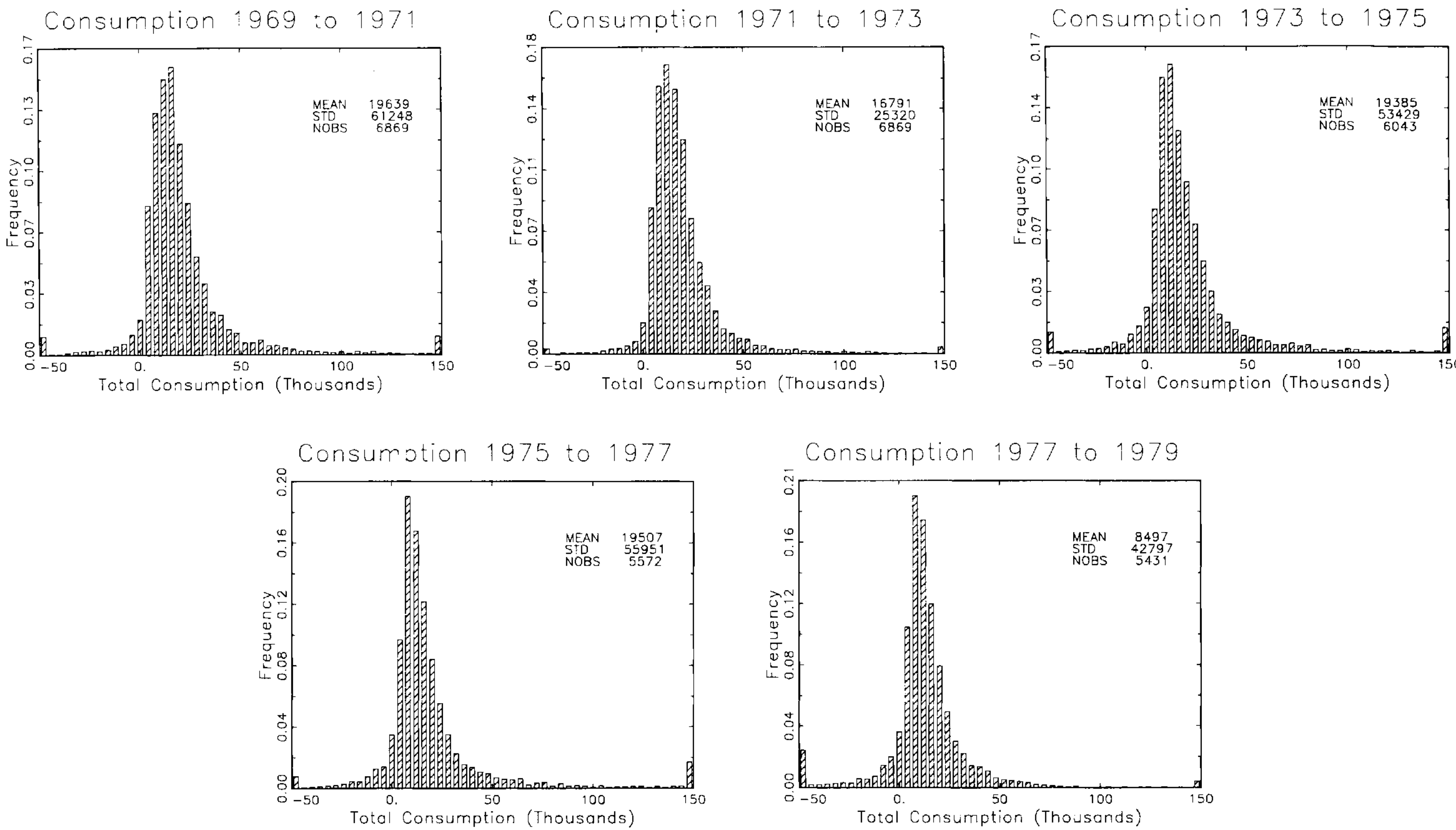

Fig. 11.7 Distribution of measured biennial consumption, 1969–79

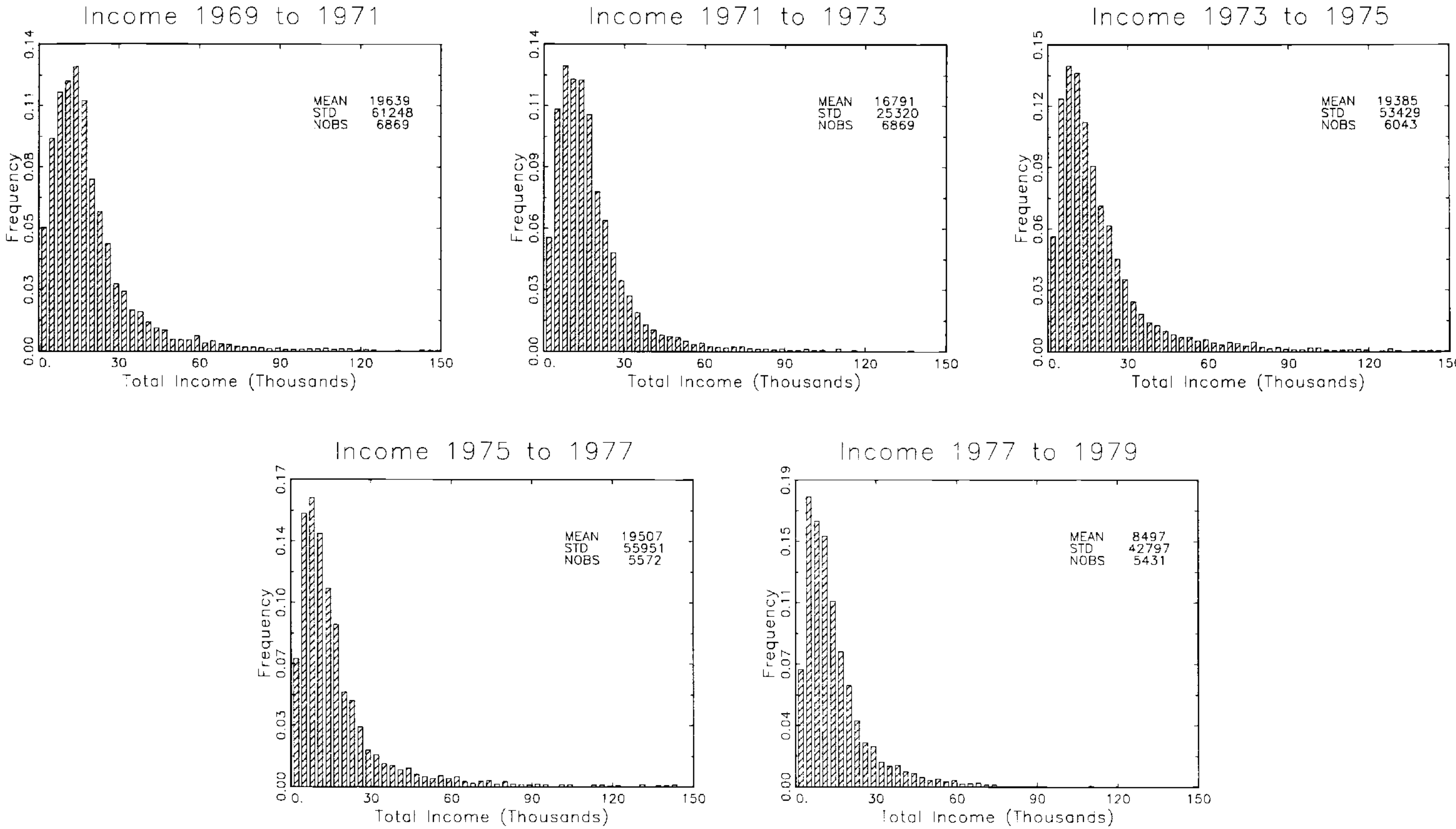

Fig. 11.8 Distribution of biennial income, 1969–79

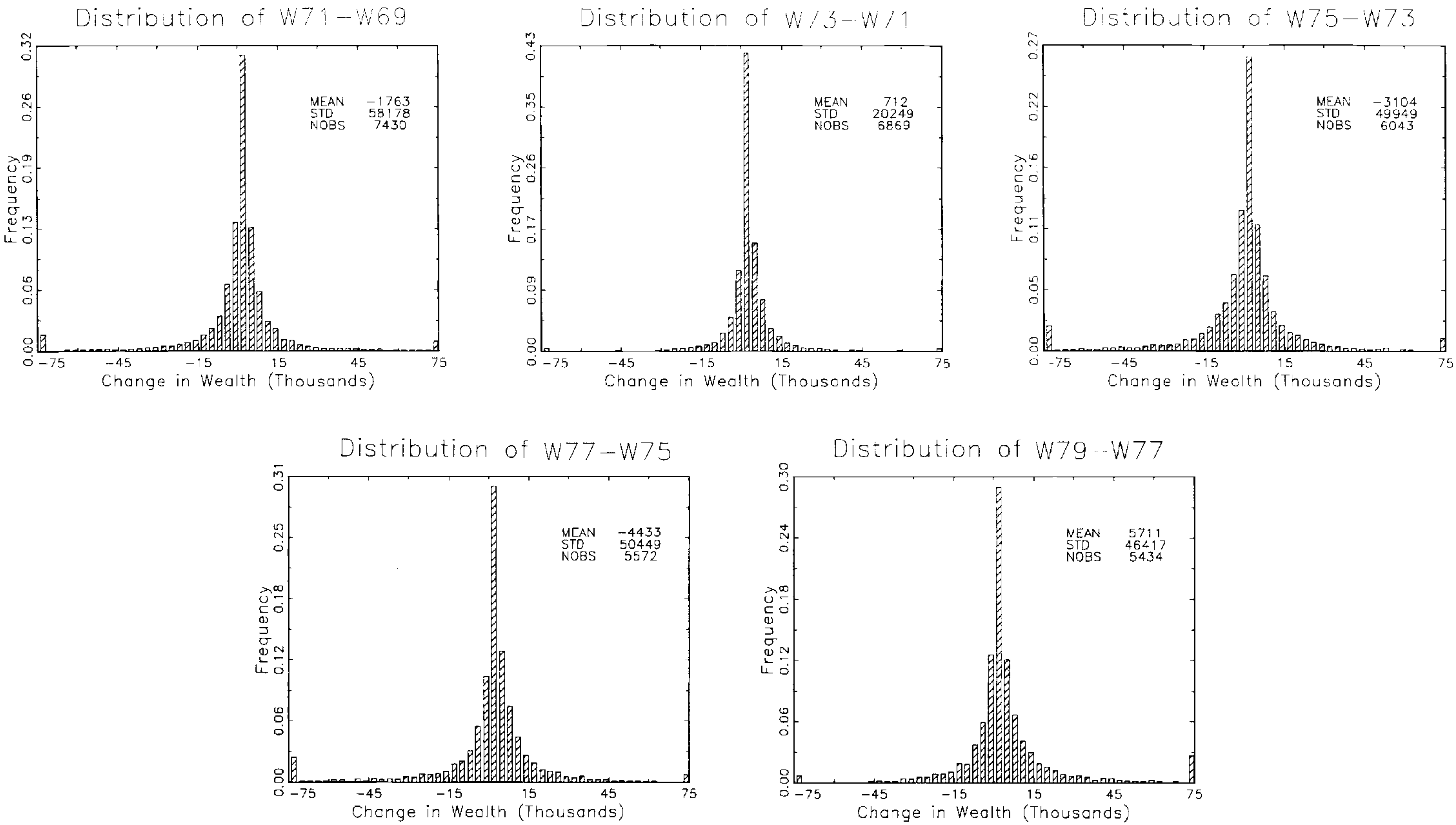

Fig. 11.9 Distribution of changes in net worth, 1969–79

Adding these housing capital gains (or losses) did reduce the number of negative consumption cases somewhat, to 1,522, but overall the distribution of consumption including capital gains looked very similar to the distribution of consumption without capital gains.

The notion that response errors in wealth are driving the violent swings in consumption is confirmed by examining individual data records. Having access to a complete data record over the survey period often provides enough contextual information to enable one intuitively to identify reporting and recording errors that are responsible for negative consumption values. Table 11.17 presents relevant data for a "typical" respondent (ID 6886) with negative measured consumption. This man—call him Bob—is coded as having the occupation of craftsman in the construction industry; most likely, Bob is a carpenter. Bob responded in all six of the survey waves and provided very complete answers; all the variable flags (with the exception of consumption) indicated very high confidence levels in his responses. Bob is married, living with spouse, and was working full time until 1975, when he turned 65, quit his job, and started collecting Social Security (Bob had no pensions). By all accounts, Bob is just the kind of guy I want in my sample: a typical blue-collar worker who seems to provide complete, reliable answers, who has slightly above average income, and who has most of his wealth in housing. However, we can see from table 11.17 that, while Bob's income declined slightly from \$12,000 to \$11,000 over the decade and his measured consumption was about equal to his income stream in four out the five two-year periods, for some reason his consumption over the period of 1976–77 is recorded as \$ − 35,630. Analysis of his balance sheet reveals that, between the 1975 and the 1977 interviews, his house value increased from \$14,000 to \$100,000, increasing his overall net worth from \$10,794 to \$57,376.[27] This sudden increase in wealth is responsible for the recorded negative consumption in the 1976–77 biennium. A possible explanation for the increase is coding error: Bob may have reported his house value to be \$10,000 in 1977, but it was mistakenly recorded as \$100,000. However, this explanation becomes less plausible when we realize that his house value is recorded at \$150,000 in 1979: it seems very unlikely that we would get the same kind of coding error in the same variable in two consecutive years.

If we look further into the data, we find that Bob moved between 1975 and 1977. This suggests several possibilities. Bob and his wife may have moved into the house of his wealthy son and mistakenly reported the value of his son's house as his own. Bob may have previously grossly underestimated the value of his old house and used the capital gains on the sale of the old house to finance the purchase of his new house. Bob may have won a lottery, which provided an unrecorded capital gain that he used to purchase his retirement dream home. Bob may have owned other real estate which he failed to report in previous interviews and has since used to buy his retirement home. Or, being a carpenter, Bob may have built his own retirement home and, blinded by the pride of creation, grossly overestimated its value. Given the wealth of possible

Table 11.17 Selected Financial Data for "Bob," (RHS ID, 6886)

	1969	1971	1973	1975	1977	1979
Personal data:						
a_t	59	61	63	65	67	69
ms_t	1	1	1	1	1	1
h_t	1	2	1	1	2	2
Employment data:						
IE	1	1	1	1	3	3
SR	1	1	1	2	3	3
SE	1	1	1	2	3	3
Financial data ($1968):						
w_t	10,698	12,523	13,950	10,794	57,376	71,555
y_t	*M*	12,033	11,951	10,199	10,952	11,429
c_t^a	*M*	10,196	10,524	13,354	−35,630	−2,750
c_t^b	*M*	11,154	12,047	12,272	−35,630	11,151
Capital gains	*M*	958	1,523	−1,082	0	13,901
Balance sheet (nominal):						
House value	8,000	10,000	13,000	14,000	100,000	150,000
Mortgage	0	0	0	0	0	0
Other house debt	0	0	0	0	0	0
Farm value	0	0	0	0	0	0
Farm mortgage	0	0	0	0	0	0
Business value	0	0	0	0	0	0
Business debt	0	0	0	0	0	0
Real-estate value	0	0	0	0	0	0
Real-estate debt	0	0	0	0	0	0
Auto value	2,490	2,490	2,495	2,500	3,237	4,980
Auto debt	0	0	0	0	0	0
Savings bonds	2,000	2,000	2,408	0	0	0
Stocks	0	0	0	0	0	0
Credit card debt	0	200	236	0	0	500
Checking account	190	900	900	900	900	1,080
Savings account	0	0	0	0	0	0
Face value life ins	1,000	1,000	913	2,000	2,000	2,000
Face value annuities	0	0	0	0	0	0
Medical debts	834	0	0	0	0	0
Store debts	100	0	0	0	0	0
Bank debts	0	0	0	0	0	0
Personal debts	0	0	0	0	0	0

[a]This measure of c_t does not include imputed capital gains.
[b]This measure includes imputed capital gains as described in the text.

explanations, it is not easy to know what to do. One can simply exclude cases with negative measured consumption, but that still leaves the problem of hundreds of cases with implausibly large or small measured consumption or cases where consumption changes vary erratically from year to year.

In conclusion, while one might attempt to identify reporting problems by examining observations on a case-by-case basis, it is unrealistic to think that one could screen out a sufficiently high fraction of "bad" cases to end up with a subsample for which consumption is measured accurately. Not only is case-by-case examination of 8,131 individuals impossibly time consuming, but the resulting data set would be susceptible to the criticism that the sample had been "hand picked" to support an *a priori* theory. If an error-identification strategy is to be successful, one should be able to write out a series of objective classification rules, say in the form of a computer program, that would allow other researchers to replicate the subsample. I have not been successful in constructing a computer program with sufficient "intelligence" to examine the wealth data on a case-by-case basis, recognize the existence of a data problem, and take appropriate corrective action. As I discussed above, it is not sufficient simply to screen out cases with negative consumption because the remaining cases still suffer from reporting problems that produce unrealistically large swings in consumption. Because of these problems, I have opted against using consumption data in my first attempts at estimating the DP model. Until I see convincing evidence that changes in wealth are not dominated by measurement error, or until I am successful in constructing an "artificial intelligence" program that reliably discriminates accurate survey responses from inaccurate responses, I will adopt the null hypothesis that $c_t = y_t$ and focus on "explaining" the joint dynamics of $x_t = (y_t, e_t, sr_t, a_t, ms_t, h_t)$ and $d_t = (s_t, ss_t)$, excluding w_t and c_t from the model.

11.8 Estimating the Stochastic Process of Income

All that remains is to specify and estimate the final component of workers' beliefs, the transition density for income π_y. The lognormal shapes of the income distributions plotted in section 11.7 suggest that the transition density π_y should have a lognormal distribution with parameters $(\mu\ \sigma)$ that are parametric functions of the state and control variables listed in the decomposition (5). As is well known, if a random variable $\tilde{y}$ has a lognormal distribution, then its mean and variance are given by

$$E[\tilde{y}] = \exp\{\mu + \sigma^2/2\}, \qquad (7)$$

$$\mathrm{var}[\tilde{y}] = \exp\{2\mu + 2\sigma^2\} - \exp\{2\mu + \sigma^2\}.$$

It is extremely important to allow both μ and σ to depend on the state variables since, if σ is fixed, then (7) and the autoregressive properties of the income process will imply that the variance of y_{t+1} is a quadratically increasing function of current income y_t. Thus, by failing to specify σ properly, one is making an implicit assumption about the form of heteroscedasticity that may grossly misrepresent workers' actual beliefs. Once we have decided on the appropriate functional forms for μ and σ, the lognormal model is fairly easy to estimate: one obtains initial estimates of (μ, σ) by a log-linear regression and uses these as starting values for computing the final parameter estimates by maximum likelihood.[28] There is a minor problem concerning the fact that the DP model requires y_t and its transition density π_y to be discretized. My approach was to discretize y_t as an independent variable entering (μ, σ) but to do the estimation treating the dependent variable y_{t+1} as a continuous variable. After estimating the relevant parameters, it is easy to generate a discrete transition probability matrix $\hat{\pi}_y$: simply compute the area under the lognormal density corresponding to each of the discrete income cells for y_{t+1}.

The hard part is to specify how the parameters (μ, σ) depend on the underlying state and control variables. The specification is crucial here because not only must π_y embody workers' expectations about how future income depends on their current employment, health, and marital status but it must also embody the relevant rules and actuarial structure of the Social Security OASDI system, including the regressive nature of the payout schedule, the extra payments to spouse, the penalty for early retirement, and the "earnings test" for workers under 70. As I discussed in my earlier paper, by estimating π_y using income data over the decade of the 1970s (during which Social Security benefits increased more than 50 percent in real terms), I have implicitly assumed that workers have "semirational" expectations: that is, they correctly anticipated the increase in benefits over the 1970s but did not expect any benefit changes thereafter.[29]

My initial attempts to estimate π_y yielded disappointing results. Although the coefficient estimates for the marital status, employment status, and search variables had reasonable signs and magnitudes, the variables representing the structure of OASDI benefits either had small, insignificant coefficients or else had the wrong sign. The estimated model looked as if workers were unaware of key features of the OASDI benefit plan, and the few provisions they did know about seemed to be regarded as taxes instead of benefits. Apparently, the Social Security benefit structure was "drowned out" by sample selection bias. A simple explanation of the problem goes as follows. High-income workers typically continue working beyond retirement age and delay collection of Social Security, whereas low-income workers stop working and begin collecting Social Security as soon as they can, typically at age 62. A regression model attempts to fit the data by flipping the sign of Social Security variables: collection of Social Security benefits is spuriously predicted to reduce total income. My solution to the problem was to augment the data set with "artificial" data on the incomes that retired workers would have received in the absence of OASDI payments. Thus, corresponding to each data record for a retired worker receiving OASDI ($ss_t \in \{1, 2\}$), I created a duplicate record deducting all OASDI benefits from the worker's income y_t and setting $ss_t = 0$. This procedure, which nearly doubled the number of observations, produced dramatically improved results. In particular, nearly all the Social Security variables had significant coefficients with correct signs and magnitudes. In effect, the augmented data "drowned out" the sample selection bias, allowing me to capture the true underlying OASDI benefit structure more accurately.

The existence of the SSMBR data set was absolutely crucial to the success of this procedure since, as I have shown, the magnitude of response error in the self-reported values of certain Social Security benefits such as SSDI is so large as to render them useless.

A final problem I encountered concerned the estimation of age-income effects. In my initial specifications, I included the polynomial terms in the age variable a_t to capture the independent effects of aging on income. Just looking at the estimated coefficients, the estimated model seemed quite reasonable, with age terms all entering with highly significant coefficients. However, when I plotted out the age-income profiles, the results were clearly far from reasonable. In models that included only a linear term in a_t, the age-income profile sloped upward, whereas in models with quadratic and cubic age terms the age-income profile was hump shaped: rising until age 70 and then falling sharply thereafter. The incomes predicted by the hump-shaped profiles were completely unreasonable: at the top of the hump a 70-year-old worker who was currently earning \$10,000 could expect to earn nearly double that amount two years later if he continued working. On the other hand, on the downward sloping part of the profile, say at age 80, the worker would only expect to make half as much even if he continued working. The reason behind these strange results is lack of data on earnings for very old men. As I have discussed before in section 11.4, the RHS has no data on workers older than 73. Thus, estimation of age-income profiles beyond age 73 requires pure extrapolation over a region where there are no observations to guide us. Including polynomial age terms in the regression produced unreasonable forecasts because the estimation procedure chose the coefficients to get a good fit in the region where there are a lot of observations, namely, for ages 58–68. Since there are no observations beyond age 73, the regression does not "care" what its predictions are in that range, producing unreasonable results. In order to avoid the extrapolation problems inherent in the use of polynomial terms, I tried specifications using age dummies, which entail the implicit extrapolation that age-income profiles are constant after age 73. In spite of my hopes, the age dummies also yielded somewhat disappointing results: the estimated age-income profile fluctuated up and down with no clear pattern. Since I have little a priori knowledge of the correct shape of the age-income profile, I decided simply not to include a_t in the estimation of π_y.

Table 11.18 presents the specification for π_y that I finally settled on. The main implications of table 11.18 have already been discussed in conclusion 6 of section 11.2 and will not be repeated here. However, to convince the reader that the estimated model really does endow workers with sensible income expectations, I present a graphic summary of the predictions of the model in figure 11.10.

Figure 11.10 presents the estimated transition densities $\hat{\pi}_y$ for four configurations of the conditioning variables listed in (5) corresponding to the beliefs of four different workers about their future income, $\tilde{y}_{t+1}$. The sharply peaked density marked with triangles represents the expectations of a single man, aged

Table 11.18 Estimates of Income Transition Probability (dependent variable: $\ln[y_{t+1}]$)

Variable	Parameter Estimates	Corrected t-Statistic
	σ Parameters	
Constant	−.25	−8.8
$\ln(y_t)$	−.51	−31.9
	μ Parameters	
Constant	−.12	−2.8
$\ln(y_t)$	.94	161.6
$h_t = 1, h_{t+1} = 1$	.02	3.8
$h_t = 1, h_{t+1} = 3$	−.27	−4.1
$h_t = 2, h_{t+1} = 2$	.01	.6
$h_t = 2, h_{t+1} = 3$	−.24	−3.5
$h_t = 3, h_{t+1} = 3$	.05	2.9
$s_t = 1, e_{t+1} = 1$	.22	14.2
$s_t = 1, e_{t+1} = 3$	−.24	−11.7
$s_t = 2, e_{t+1} = 1$	.19	7.4
$s_t = 2, e_{t+1} = 2$	.00	.1
$s_t = 2, e_{t+1} = 3$	−.31	−11.9
$s_t = 3, e_{t+1} = 1$	.02	.3
$s_t = 3, e_{t+1} = 2$	−.03	−1.2
$s_t = 3, e_{t+1} = 3$	−.18	−10.7
$s_t = 3, e_{t+1} = 3, y_t < 4$	−.08	−7.1
$s_t = 3, e_{t+1} = 3, y_t > 15$	−.11	−3.8
$ms_t = 2, ms_{t+1} = 2$	−.19	−4.7
$ms_t = 1, ms_{t+1} = 2$	−.30	−6.2
$ms_t = 1, ms_{t+1} = 1$	−.04	−1.2
$ss_t \neq 0, ms_t = 2, ms_{t+1} = 2$[a]	−.04	−.9
$ss_t \neq 0, e_{t+1} = 1$[a]	.04	1.8
$ss_t \neq 0, e_{t+1} = 2$[a]	.47	10.7
$ss_t \neq 0, e_{t+1} = 3$[a]	.52	23.1
$ss_t = 2, e_{t+1} = 1$[a]	−.01	−.4
$ss_t = 2, e_{t+1} = 2$[a]	.01	.2
$ss_t = 2, e_{t+1} = 3$[a]	.19	7.9
$ss_t \neq 0, a_t \geq 70, e_{t+1} = 1$[a]	.36	2.3
$ss_t \neq 0, a_t \geq 70, e_{t+1} = 2$[a]	.33	3.8
Log likelihood	−2.9 E+5	
Grad • direc	2 E-026	
Number of observations	39,494	

[a]These variables are all multiplied by $1/\ln(y_t)$.

75, who is disabled, out of the labor force, and receiving a total income of y_t = \$4,000. The density marked with the circles corresponds to a 65-year-old retired man who is married, in health state $h_t = 2$, and receiving an income of y_t = \$7,000. The density marked with boxes corresponds to a married 58-year-old man who is in good health, working full time, with a total income

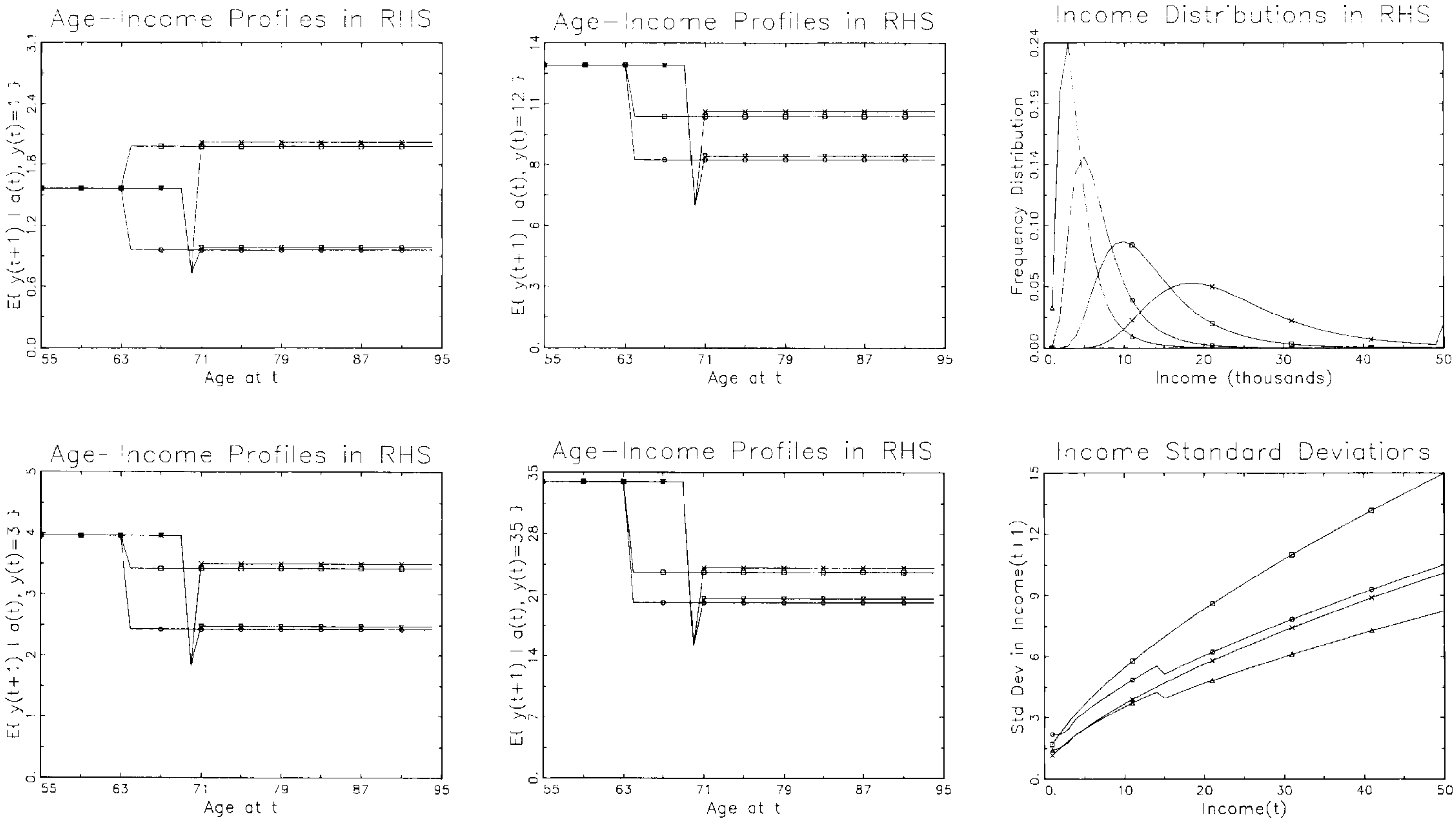

Fig. 11.10 Beliefs about future earnings implicit in estimated income transition probabilities

of y_t = \$12,000. Finally, the curve marked with x's corresponds to a wealthier 80-year-old man who is not retired, married, in good health, and continues to work full time, earning an income of y_t = \$20,000.

Figure 11.10 also presents estimates of workers' expected income, $E\{y_{t+1}|e_{t+1}, ms_{t+1}, h_{t+1}, x_t, d_t, a_t\}$, plotted as a function of their age. Although the profiles are flat by construction (the estimated model excluded a_t), the figures provide an indication of the dynamics of income as workers retire. Each of the figures contains four curves, corresponding to four different retirement paths. The curves marked with boxes correspond to working full time until age 65 and then collecting OASI. The curve marked with circles shows what the worker would expect if he quit working but did not start collecting OASI. The other two curves represent the expectations of a worker who works full-time until his early 70s but then becomes disabled and has to quit work. The lower curve represents what the worker would expect if there were no OASDI program to cover him; the higher curve represents what the worker would expect if he applied for OASDI. Note carefully that the curves in figure 11.10 represent conditional expectation functions: they are not the same as the sample paths of the income process. Given the strong autocorrelation in income, actual sample paths of income will look quite a bit different. The figure clearly shows the progressive nature of the Social Security system. Indeed, a very low-income worker actually expects to do better by retiring and collecting OASDI than continuing to work at his low-paying full-time job. However, figure 11.10 shows that, for a very high-income worker, the percentage replacement rate of OASDI benefits is much smaller: Social Security is not such a good deal for these workers.

Finally, figure 11.10 also includes a plot of the standard deviation of $\tilde{y}_{t+1}$ as a function of current income, y_t. The four curves are all upward sloping, representing the fact that, the higher a worker's current income is, the more uncertain he is about his future income. Note that, while uncertainty does increase with y_t, the increase is far from proportional: this is a direct consequence of the fact that $\ln(y_t)$ enters the σ parameter with a large, significant negative coefficient, as you can see from table 11.16. The four curves in the figure correspond to four classes of workers. The curve marked with boxes corresponds to a 60-year-old worker who is married, in good health, and working full time. The curve marked with circles corresponds to a worker who is 88, disabled, and out of the labor force. The curve marked with triangles corresponds to a worker who is 68, single, in health state $h_t = 2$, and is retired and receiving Social Security. The final curve, marked with x's, corresponds to a 55-year-old man who is single, in health state $h_t = 2$, and working part time.

11.9 Modeling the Retirement Decision

The SSMBR data allow me to determine exactly when a worker applys for and receives OASI benefits. In my opinion, the only sensible and precise treatment of the concept of "retirement" is to define it in terms of collection of OASI benefits. I used the SSMBR data set to construct the control variable ss_t defined in (2). Figure 11.11 summarizes this variable in terms of the implied distribution of age of first entitlement to OASDI.[30] The distribution has a pronounced bimodal shape, with peaks at the early retirement age 62 and at the normal retirement age 65. Overall, nearly all the sample applies for benefits between the ages of 62 and 65.[31]

One can, however, define retirement in terms of withdrawal from the labor force. The discussion in section 11.5 indicates that this is an extremely tricky business since there is no clear-cut way to define "withdrawal from the labor force." Figure 11.11 presents a "retirement age" distribution tabulated by Burtless and Moffitt (1984), who used the RHS data, the instantaneous measure of labor participation IE, and a definition of "retirement" to be a sudden, discontinous drop in labor supply to under thirty hours per week. This distribution is significantly more spread out than the distribution of age of entitlement to OASDI. In particular, the peaks at ages 62 and 65 are much less pronounced, and there are much larger fractions of workers retiring before and after ages 62 and 65, respectively. Figure 11.11 also presents similar distributions tabulated from the RHS by Sueyoshi (1986), using instantaneous hours of work data and still another definition of retirement, and finally my own tabulation based on my classification of the workers' employment history (i.e., for workers' following the standard "1–2–3" and "1–3" employment sequences, retirement is defined as the age at which the worker first enters employment state 2 or 3; for those with nonmonotonic employment sequences, it is defined as the age at which the worker first begins collecting OASDI). Although each of the definitions yields a significantly different retirement age distribution, it is difficult to say which is the "right" one. However, it turns out that both the Burtless and Moffitt and the Sueyoshi distributions significantly understate the number of early age 62 retirements and overstate the number of age 65 retirements. While it is likely that their definitions of retirement have obscured some important features of the data, this analysis suggests that a debate about the "correct" definition of retirement age is simply ill posed in the context of a more realistic dynamic model of employment transitions.

What we can conclude from figure 11.11 is that there are a significant number of workers who apply for OASI in their mid-60s but continue working for several more years. One can see this most clearly by comparing the distribution of the age at which respondents were first entitled to OASDI versus the distribution of ages at which they first received six or more months of OASDI benefits (an indirect indicator of withdrawal from the labor force). The peak of early retirements at age 62 is nearly identical in both graphs: the primary differences are a shift in probability mass from retirements in the 63–65 age group to the age 66–72 age group and a near doubling of the number of respondents who never ended up collecting OASI benefits at all, from 687 to

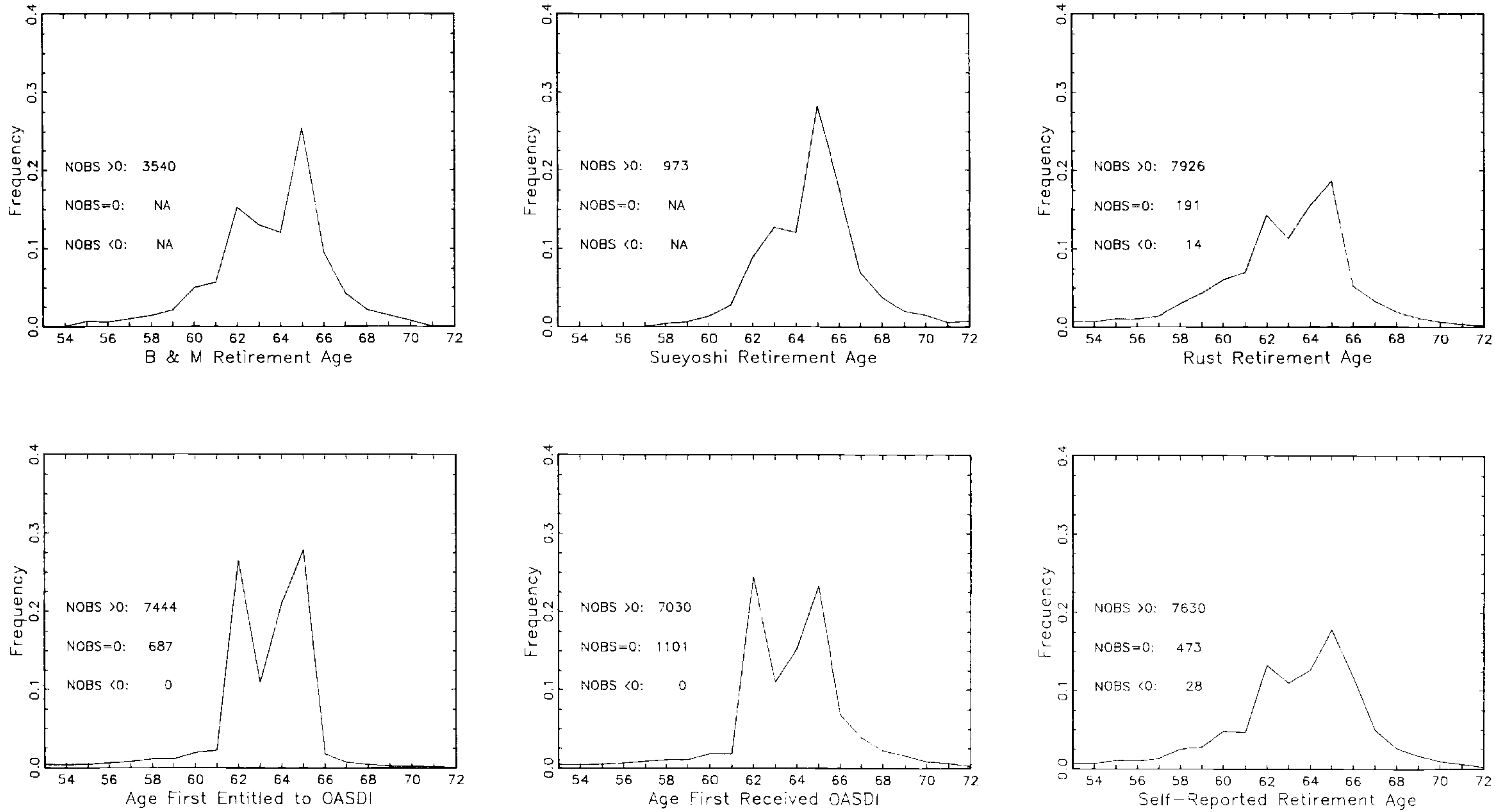

Fig. 11.11 Distributions of retirement ages implied by alternative definitions of "retirement"

1,101. There are four main reasons why the subsample of 1,100 workers never collected OASDI benefits: (1) the worker died before he had a chance to receive benefits (458 cases); (2) the worker was not eligible for OASDI, either due to the fact that he was not in a covered occupation or had not accumulated sufficient quarters of coverage to qualify for benefits (405 cases); (3) the worker applied and was entitled to OASI but lost 100 percent of his benefits owing to earnings well in excess of the Social Security earnings test (184 cases), and (4) the worker never applied for benefits even though he was eligible (fifty-four cases). Nearly all workers in the latter group had sufficiently high earnings that they would not have received benefits even if they had applied. Only four or five cases can be identified where an unemployed worker apparently "forgot" to apply for his Social Security benefits. Among the subsample of workers who ultimately collected OASI benefits, there is a two-year average delay between application and first receipt, from age 64 to 66, primarily due to that fact that benefits were taxed away due to earnings in excess of the earnings test levels. However, the predominant majority of this sample, 71 percent, first collected benefits in the same year that they became eligible. Approximately 18 percent began collecting one year after their entitlement, another 6 percent began collecting two years after entitlement, and the remaining 5 percent collected benefits three to nine years after first entitlement. This raises the question, Is it irrational to apply for benefits before age 65 yet continue working up until the normal retirement age and thereby lose nearly all benefits due to the earnings test? Close inspection of the Social Security regulations reveals that this is not irrational behavior. Section 729 of the 1974 *Social Security Handbook* provides for *ex post* adjustment of the early retirement benefit reduction factor to exclude months between ages 62 and 65 for which work deductions were imposed. Thus, even though a worker has applied for benefits before age 65, his ultimate reduction factor is based on his age when benefits are first paid. Early application simply provides an option for immediate collection of benefits in the event of unexpected unemployment but does not necessarily imply a permanent reduction in benefit levels.

Further insight into the issue is provided by the distribution of "self-reported" retirement age, defined as the age at which the worker first reports being "retired or partly retired." This distribution supports the view that, even though a large number of workers apply for OASI at age 62, many of them continue working for several years thereafter. Further analysis (beyond the scope of this paper) indicates that the majority of workers who apply for benefits at age 62 and continue working are either low-wage/income workers whose total annual earnings at ages after 62 are not significantly higher than the earnings test level or are a smaller group of workers who apparently initially intended to quit working at age 62 but experienced adverse financial problems or encountered a particularly attractive job opportunity that prompted them to return to work. But by far the biggest discrepancy between the entitlement and labor force withdrawal definitions of retirement is for a

10 percent subsample of wealthy professionals and self-employed workers who apply for Social Security benefits by age 65 but continue working well into their 70s. This type of behavior is evidently not irrational.

By in large, the analysis and estimation results in the previous sections of the paper suggest two main conclusions about the decision to retire and collect OASI: workers who retire and receive OASI appear to be less healthy than their counterparts who continue to work; and, once a worker starts collecting OASI, he is significantly less likely to return to work on either a full- or a part-time basis.

Finally, analysis of the RHS data provides clear evidence of the role of self-selection in the decision to collect OASDI; poorer, less healthy workers are more likely to quit their jobs and retire early, even given the permanent 20 percent penalty for early retirement. This may be rational behavior given the reduced life expectancy of unhealthy workers and the well-known fact that the OASI benefit structure is not actuarially fair. A more complete analysis of these issues must await the estimation of the DP model in the third part of this series.

Notes

1. For example, there is no attempt to measure consumption flows from durable goods such as housing, automobiles, furniture, etc. The questionnaire also requires estimates for expense categories that may be very hard to recall, e.g., amount spent in restaurants, amount spent for newspapers, amount spent for haircuts, etc. In fairness, I should mention that some authors, such as Hurd (1990), have attempted to use this data in an attempt directly to impute total consumption c_t. Other authors, such as Skinner (1987), propose using explanatory variables common to the more complete Consumer Expenditure Survey to compute regression-based imputations of consumption.

2. The wealth and income data used in this study were produced by the program IMPUTE, written by Beth van Zimmerman and Phil Farrell, research associates of Michael Hurd, State University of New York at Stony Brook. Besides imputing missing values, the program estimated the value of service flows for owned assets such as autos and housing at a presumed opportunity cost of 3 percent.

3. Constructing labor force histories turned out to be a major undertaking, requiring over eighty pages of FORTRAN code and over four months of full-time work to write and debug. The difficulties arose from the need carefully to track the survey skip patterns to extract the required variables from a battery of more than 130 questions in the "Work Experience" section of the RHS. Fine attention to detail was required to avoid misclassifying 20 percent of the sample of workers with "nonstandard" employment histories with multiple job transitions within the two-year survey period.

4. Of course, I will be happy to provide the reader with the data and documented versions of all computer programs used to generate the variables so that other researchers can verify any of my results, should they choose to do so.

5. Bienniel income was used only for purposes of constructing a measure of consumption. Based on conclusions 2 and 3 above, I have decided to exclude

consumption/savings decisions and formulate a DP model with biennial time intervals, measuring workers' states over the preceding even-numbered survey years. Thus, the DP model will actually use the annual income flows that were recorded in the surveys. For further justification of this approach, see conclusion 6.

6. A related problem with the disability classification, the fact that the probability of becoming disabled declines sharply to zero at age 62, is also an artifact of Social Security rules and is discussed further in conclusion 2 of sec. 11.2.

7. This decision is relevant only for workers who are over 62, have sufficient quarters of coverage to qualify for fully insured status, and have not previously applied for Social Security.

8. Cross-validation procedures can be used to fix values for some of the smoothing parameters, but ultimately many choices, such as the functional form of the kernel, are completely arbitrary.

9. The latter conclusion disappears if I exclude the variable sr_t distinguishing respondents who are receiving OASDI. Since men who collect OASDI have higher death rates, excluding sr_t produces a model where death rates increase slightly with age.

10. Social Security law prohibits a disabled worker from engaging in "substantial gainful activity" unless they are over age 62, at which time benefits convert to retirement benefits and are subject to the usual earnings test. Disabled workers are allowed to participate in a nine-month trial work program, after which continued work leads to termination of disability benefits. However, very few disabled workers ever return to work. Modeling the underlying decision process of whether to apply for disability benefits is hampered by lack of data on respondents who applied for and were denied disability benefits.

11. The finding for single workers might be partly a result of sample selection bias: single workers are presumably less likely to have a family support network to rely on, so they are more likely to become institutionalized if they have serious health problems. Such workers are lost from the sample since the RHS did not attempt to interview institutionalized individuals.

12. Here, s_t is proxied by the NE measure of the employment search decision defined in section 11.1.

13. Recall that my income measure includes net asset income (including imputed income from assets such as net housing equity), non–Social Security pension income (although the sample selected out individuals who had significant pension income), and income from relatives and other sources. My estimates of the reduction in income from retirement (defined here as quitting the labor force and applying for Social Security benefits) are somewhat higher than those of Fox (1984), who found that retirement leads to a 30 percent median percentage drop in total income for workers without private pensions.

14. In order to keep the length of this paper within bounds, I have chosen not to present the estimation results that lead to this conclusion. I defer the presentation of the results to the third paper of the series, which will examine the heterogeneity issue in more detail.

15. The nonresponse file was compiled by the Census in the process of conducting the RHS interviews and was used by SSA as part of an internal auditing system to remove cases in which the interviewer was unable to contact the original 1969 respondent or related household members. For some reason, the nonresponse data were not included on the RHS tapes and are available only separately as a subfile of the SSMBR tape. The nonresponse file will also be used in sec. 11.4 to identify men who were institutionalized after the 1969 interview.

16. Note that the parameter standard errors and t-statistics have been corrected using White's (1982) formula.

17. Although the total number of cases seems small, think of the millions of unnecessary tax dollars spent if this error rate exists in the population at large.

18. The aggregate mortality rates were obtained from the *Statistical Abstract of the United States* (U.S. Census 1979). In future work, I would like to formally incorporate auxiliary mortality data for very old men into a pooled maximum likelihood estimation of the death hazard model. A difficulty of this approach is the likely absence of associated health and employment status in any auxiliary data set. This will require me to "integrate out" these variables, which in turn requires further distributional assumptions on the cross-sectional distributions of health and employment status for very old men. Given these problems, I decided to use the short-cut described above.

19. I collapse those sequences for which M's occur only as trailing sequences. Cases where there are intervening occurences of M's (respondents who missed one survey but were subsequently interviewed) are classified in the "others" category in table 11.9.

20. The NLS contained an enriched sample of black respondents, who are presumably more likely to be unemployed. Apparently, the effect of a more youthful sample in the NLS dominated the effect of a larger proportion of blacks, leading to the discrepancies noted above.

21. Since probabilities sum to one, it is not necessary to present parameter estimates for the event $I\{e_{t+1} = 2\}$, which is equivalent to normalizing the parameters corresponding to the event $I\{e_{t+1} = 2\}$ to zero.

22. I substituted actual OASDI benefits from the SSMBR rather than reported OASDI benefits to calculate total income in even-numbered years.

23. Workers whose imputed wage appeared to be either unreasonably large or small or whose wage rate changed significantly were flagged and reexamined. Many of the unreasonable cases appear to be a consequence of reporting errors in income or employment status. The income distributions presented in figures 11.3, 11.4, and 11.8 do not screen out these questionable cases, however.

24. The "actual" wage income was taken as the value recorded by Social Security in the SSER data set. This income measure is right censored at the Social Security maximum earnings levels of $7,800 in effect over the period 1968–71. I have not attempted to use the quarters of coverage data to impute actual total wage earnings according to the method of Fox (1976).

25. While pensions are much more common than annuities in the RHS sample, exclusion of pension wealth is not a problem since the sample boolean already excludes workers with pensions.

26. The hypothesis actually tested was H_0: $w_t = w_{t+1}$ vs. H_A: $w_t \neq w_{t+1}$. It is easy to see that the budget identity implies that this is equivalent to the hypothesis test listed above. I assume that appropriate regularity conditions hold in order to justify the asymptotic χ^2 distribution for the test statistic, e.g., weak mixing conditions on the level of serial dependence in the observations.

27. Note that no capital gains are imputed since Bob moved during the period, making it impossible to determine how much of the value of the new house came from capital gains on the sale of the old house.

28. Although the likelihood function is concave, using the regression starting values (as opposed to zero starting values) substantially reduces the number of iterations needed to converge.

29. In fact, the large benefit increases in the 1970s put severe strain on the Social Security trust fund, necessitating substantial tax increases to fund the system. Fully rational workers might have plausibly expected real benefit *decreases* in the future.

30. A worker is entitled to OASDI if he (1) is at least age 62, (2) has filed a valid application for benefits, and (3) has sufficient quarters of coverage to qualify for retirement benefits. If the worker is under age 62, he is entitled to disability benefits

if he filed a valid application and was granted disability insured status by the Social Security Administration.

31. The fact that so few workers apply for benefits after age 65 might initially appear somewhat surprising in light of the fact that a significant number of higher-income workers continue working full time well into their 70s. It turns out that, under Social Security rules, there is no reason to delay application for benefits after age 65 since workers must first be entitled to OASI in order to be entitled to Medicare hospital insurance coverage (cf. sec. 104-A of the 1974 *Social Security Handbook*). Since Social Security benefits are automatically recomputed after initial entitlement, there is no benefit penalty (owing to the computation of average monthly wage) to delaying application beyond age 65. In addition, delayed retirement credits are determined on the basis of the worker's age when he is *first paid* benefits rather than on the worker's age when he applied. All these factors of the benefit structure explain why hardly anyone applies for benefits after age 65. I also take it as *prima facie* evidence that nearly all workers are cognizant of the Social Security benefit regulations.

References

Berkovec, J., and S. Stern. 1987. Job exit behavior of older men. University of Virginia. Typescript.

Bertsekas, D., and S. Shreve. 1978. *Stochastic optimal control: The discrete time case.* New York: Academic.

Bound, J. 1986. The disincentive effects of the Social Security disability program. University of Michigan. Typescript.

Burtless, G., and R. Moffitt. 1984. The effect of Social Security benefits on the labor supply of the aged. In *Retirement and economic behavior,* ed. H. J. Aaron and G. Burtless. Washington, D.C.: Brookings.

Fox, A. 1976. Alternative measures of earning replacement for Social Security benefits. Research Report no. 47. Office of Research and Statistics, Social Security Administration.

______. 1979. Earnings replacement rates for retired couples: Finding from the Retirement History Study. *Social Security Bulletin* 42:17–39.

______. 1982. Earnings replacement rates and total income: Findings from the Retirement History Study. *Social Security Bulletin* 45:3–24.

______. 1984. Income changes at and after Social Security benefit receipt: Evidence from the Retirement History Study. *Social Security Bulletin* 47:3–23.

Gustman, A. L., and T. L. Steinmeier. 1983. Minimum hours constraints and retirement behavior. *Economic Inquiry* 3:77–91.

______. 1984. Partial retirement and the analysis of retirement behavior. *Industrial and Labor Relations Review* 37:403–15.

______. 1986. A structural retirement model. *Econometrica* 54(3):555–84.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–54.

Hill, D. 1988. Response error around the seam: An analysis of change in a panel with overlapping reference periods. In *Individuals and families in transition: Understanding change through longitudinal data,* 277–88. U.S. Bureau of Census compilation of papers presented at the 1988 Social Science Research Council, Annapolis, Maryland.

Hurd, M. 1990. Consumption, wealth, and bequests. Typescript. State Univ. of New York, Stony Brook.
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): 265–89.
Mott, F. L., and R. J. Haurin. 1985. Factors affecting mortality in the years surrounding retirement. In *Retirement among American men,* ed. H. Parnes. Lexington, Mass.: Lexington.
Parnes, H. S., ed. 1981. *Work and retirement.* Cambridge, Mass.: MIT Press.
Parnes, H. S., et al. 1985. *Retirement among American men.* Lexington, Mass.: Lexington.
Rust, J. 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55 (5):999–1033.
______. 1989. A dynamic programming model of retirement behavior. In *The economics of aging,* ed. D. Wise, 359–98. Chicago: University of Chicago Press.
______. 1988. Maximum likelihood estimation of discrete control processes. *SIAM Journal on Control and Optimization* 26(5):1–19.
Skinner, J. 1987. A superior measure of consumption from the Panel Study of Income Dynamics. *Economics Letters* 23:213–16.
Stock, J. H., and D. A. Wise. 1989. Pensions, the option value of work, and retirement. Kennedy School of Government, Harvard University. Typescript.
Sueyoshi, G. 1986. Social Security and the determinants of full and partial retirement: A competing risks analysis. University of California, San Diego. Typescript.
U.S. Census. 1979. *Statistical abstract of the United States.* Washington, D.C.: U.S. Government Printing Office.
van Dijk, N. M. 1984. *Controlled Markov processes: Time discretization.* CWI Tract II. Amsterdam: Mathematische Centrum.
White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–26.

Comment Angus Deaton

This paper is a remarkable member of a remarkable sequence. Following the example of Trollope and Dickens, Rust is telling us his story in installments, each of which ends suspensefully; the next in the series will surely reveal all and resolve the tension. The mystery here is whether the trick can be done at all. Will the Retirement History Survey (RHS) yield to the calculus of stochastic dynamic programming and reveal the true story of aging and retirement? Or has structural estimation in econometrics at last attempted too much, even structural estimation in the hands of John Rust and the supercomputer? We shall have to wait another year to find out, but the latest installment, as it should, certainly serves to complicate the plot. It also

Angus Deaton is professor of economics and international affairs at the Woodrow Wilson School, Princeton University, and a research associate of the National Bureau of Economic Research.

provides a peculiarly inappropriate position from which to review the research; anything said now is likely to look foolish at the next round.

The most difficult task in applied econometrics is to make a clean transition from theory to implementation. I am referring not to the estimation or the interpretation of results but to the intermediate stage where sharply delineated theoretical notions have to be matched up to imperfect and error-ridden data. In the face of complex reality, theoretical concepts quickly lose their sharpness (what is retirement?), and even the best surveys turn out to have omitted the simplest and most important questions. This paper, which is the second in the sequence, is concerned with cleaning data and matching it to the demands of the theory. Rust's model is one of *discrete* choice, where controls are set at a limited number of positions so as to affect outcome states, each of which must also be discretely defined. The constraints facing optimizing consumers are the laws of motion of the system, the transition probabilities that govern the evolution of the state variables, conditional on their own past values and on the values of the controls. Apart from the data exercises, the specification and estimation of these probabilities is the main task of the paper.

One of the benefits of a series of papers is that it is possible to detail much of the important material that would typically be suppressed in journals. Rust has done an extraordinarily good job of laying out exactly what he has done; honesty and care shine out from every page. This paper and its predecessor are the best counterexamples I know to the accusation that high-tech econometricians care little about their data (or that those who care about numbers know nothing else!). The amount of work that has gone into data preparation is astonishing; the RHS tapes have been matched to the census, to the SSA records, and to the results of earlier researchers' work. Each observation has been multiply "flagged" to indicate assessments of data reliability and every detail of the process encoded and preserved to enable replication by other scholars. When the next installment comes and we find out whether a DP model can fit the RHS, it will be impossible to ascribe failure to inadequate data preparation or to suspect data "cleaning" for results that look too good.

But commentators are not supposed to express unbounded admiration, or at least not *only* unbounded admiration. So I should like to identify a few points in the paper where I was left feeling uneasy, where the compromises that had to be made seemed to be beginning to threaten the structure. There are two issues that I should like to draw out, both concerning methodological questions about this sort of structural modeling, as opposed to the more ad hoc or reduced-form analyses that has been adopted by most other researchers. The first concerns data quality. There are many points at which an important theoretical variable has to be replaced by a *very* imperfect substitute. There are also many magnitudes that appear to contain egregious errors of measurement, errors that are more than usually exposed through Rust's tireless analysis of the data. As Rust notes, the model that will eventually be estimated is (to say the least) nonlinear, and there is no good way of handling measurement errors in such a model. So it is going to take some extraordinarily sensitive testing of the final model to try to separate those parts of the results that are due to measurement error and those that are credible and robust.

The second issue concerns Rust's modeling of the "laws of motion of the system," the transition probabilities that govern the consumer's progress from state to state. Ideally, these laws of motion would be given, like the laws of physics, they could be set up as constraints, and the calculation and estimation of the DP could begin. But, of course, the transition probabilities are not known and have to be estimated. Rust summarizes his findings in section 11.2. Given the discretization of the states, the transition probability matrix contains some 130,000,000 elements, all of which have to be estimated from the data. Since most of the transitions are never actually observed in the sample, nonparametric estimation is not possible. Instead, Rust imposes structure on the probabilities, equation (5), and then estimates a system of logits. The structure used is a recursive one; health status is influenced only by lagged variables, current health status and lagged variables influence marital status, health and marital status jointly condition employment, while health, marital status, and employment condition the evolution of income. All this is perfectly reasonable, but of course it may not be correct. Rust tells us that, having tried various decompositions of the transition probabilities, the one that he found most plausible was the one discussed above. It is hard not to be reminded of the analogies with structural and reduced-form debate in macroeconomic modeling. It was exactly this use of "reasonable" (but arbitrary) exclusion restrictions that had much to do with the retreat toward less structured approaches, and I felt uncomfortable finding the same sort of issues in the current context. The next stage of the research, the calculation of a maximizing strategy in the face of the constraints, will ruthlessly expose any flaws in the modeling at this stage. If there is some cheap but absurd method of generating utility, the algorithm will find it. Again, there is analogy with macro, where the first wave of enthusiasm for optimal control of Keynesian econometric models quickly foundered on those models' lack of a supply side; optimal plans clearly involved eating the capital stock. Rust is too good an economist to fall into any of these obvious traps, but my feeling of discomfort remains. Are the estimates of the transition probabilities really soundly enough grounded to support the very great strain that is about to be placed on them?

There are some specific points in the paper where concern about the two general issues comes to a head. First, the consumption data, imputed from income and wealth changes, are too dreadful to use, so that it is going to be necessary, at least at first, to estimate a model that maximizes, not the expected utility of consumption, but the expected utility of income. This may not do too much harm, but it is a pity that so many of the important issues (life-cycle saving, wealth accumulation or decumulation, saving and retirement dates) are thereby lost from the analysis. Rust discounts the possibility of using the partial consumption data in the survey, but the decision might be worth some further consideration.

My second point concerns the definition of one of the two control variables, search effort in the labor market. Rust considers three possible measures: a self-reported intended hours next year; the actual hours two years hence, that is, in the next actual sample period; and the actual hours in the intermediate year. The first is rejected because "talk is cheap" and because the quantity bears little relation to outcomes. The second is rejected because it implies complete control. Fine, but I find it hard to see why the last measure is likely to be a good proxy either. While it is true that it predicts employment quite well, there are any number of reasons why it should, not all of which are consistent with it being a good proxy for search effort. Indeed, a worker might record a large number of hours, though fully intending to be unemployed next period, and a partially or totally "retired" worker could easily be searching very actively. Some workers may even take early retirement in order to search more actively for a suitable subsequent occupation.

My third, and final, point is again on the estimation of the transition probabilities, and again it is an issue of which Rust is fully aware. Heterogeneity of workers, or equivalently incomplete accounting of states, is quite likely to lead to inconsistent estimates or to spurious identification of causality. There are obvious examples throughout the paper, many of which are dealt with, such as the fact that "collecting OASI can be hazardous to your health." I was also amused to discover that, once single, a worker has less than a 7 percent chance of finding a new mate and that bachelors, who have no prospect of remarriage, can expect their income to fall at 20 percent a year. But, more seriously, I would expect heterogeneity to be a serious problem, and, although Rust evidently worries about it less, I should have been happier had some of the supporting evidence been included in the paper. Perhaps in the next installment.