

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: The Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90295-1

Volume URL: <http://www.nber.org/books/wise89-1>

Conference Date: March 19-22, 1987

Publication Date: 1989

Chapter Title: A Dynamic Programming Model of Retirement Behavior

Chapter Author: John P. Rust

Chapter URL: <http://www.nber.org/chapters/c11588>

Chapter pages in book: (p. 359 - 404)

A Dynamic Programming Model of Retirement Behavior

John Rust

12.1 Introduction

This paper derives a model of the retirement behavior of older male workers from the solution to a stochastic dynamic programming problem. The worker's objective is to maximize expected discounted utility over his remaining lifetime. At each time period t the worker chooses control variables (c_t, d_t) where c_t denotes the level of consumption expenditures and d_t denotes the decision whether to work full-time, part-time, or to exit the labor force. The model accounts for the sequential nature of the retirement decision problem and the role of expectations of the uncertain future values of state variables (x_t) such as the worker's future lifespan, health status, marital or family status, employment status, and earnings from employment, assets, Social Security retirement, disability, and Medicare payments. Given specific assumptions about workers' preferences and expectations, the model generates a predicted stochastic process for the variables $\{c_t, d_t, x_t\}$. This paper, however, focuses on the inverse or "revealed preference" problem: given data on $\{c_t, d_t, x_t\}$, how can one go backward and "uncover" the worker's underlying preferences and expectations?

One can formalize the revealed preference problem as a problem of statistical inference. The null hypothesis is that the data $\{c_t, d_t, x_t\}$ are

John P. Rust is an Associate Professor of Economics at the University of Wisconsin-Madison and a Faculty Research Fellow of the National Bureau of Economic Research.

This research is part of a project on the economics of aging funded by National Institute for Aging grant 3-PO1-AG05842-01. CPU time on the Cray-2 supercomputer was provided under grant SES-8419570 of the National Science Foundation.

realizations of a controlled stochastic process generated from the solution to a stochastic dynamic programming problem with utility function u and a stochastic law of motion π that depend on a vector of unknown parameters θ . The underlying preferences u and expectations π are "uncovered" by finding the parameter vector $\hat{\theta}$ that maximizes the likelihood function for the sample of data. Standard likelihood ratio, Lagrange multiplier, and chi-square goodness-of-fit statistics allow one to test whether or not workers are rational in the sense of acting "as if" they were solving the specified dynamic programming problem. If the data appear to be consistent with the dynamic programming model, the estimated model can be used to forecast the effect of policy changes such as reductions in Social Security retirement or disability benefits. Policy forecasts require a "structural" approach that attempts to uncover the underlying preferences u rather than the traditional "reduced-form" approach which can be viewed as uncovering the historical stochastic process for $\{c_t, d_t, x_t\}$. The problem with reduced-form methods, noted by Marschak (1953) and later by Lucas (1976), is that policy changes cause workers to reoptimize, yielding a new controlled stochastic process for $\{c_t, d_t, x_t\}$ that is generally different from the historical process of the previous policy regime. The structural approach allows one to solve the dynamic programming problem under the new policy regime and to derive a predicted stochastic process for $\{c_t, d_t, x_t\}$. Recovering the underlying utility function is also useful for quantifying the extent to which workers are hurt by various policy changes.

Unfortunately, stochastic dynamic programming problems generally have no tractable analytic solutions and are typically only described recursively via Bellman's "principle of optimality". Without such a solution it appears impossible to write down a simple, analytic likelihood function for the data. This problem may have deterred previous researchers from estimating structural models of retirement behavior that capture both uncertainty and the sequential nature of the decision process.¹ Recently, the advent of new estimation algorithms and powerful supercomputers has begun to make estimation of more realistic stochastic dynamic programming models feasible, even though such models have no analytic solution. The basic idea is very simple: the dynamic programming problem and associated likelihood function can be numerically computed in a subroutine of a standard nonlinear maximum likelihood algorithm. Rust (1988) developed a nested fixed point (NFXP) algorithm that computes maximum likelihood estimates of structural parameters of discrete control processes, a class of Markovian decision processes for which the control is restricted to a finite set of alternatives. As its name implies, the NFXP algorithm works by converting the dynamic programming problem into the problem of computing a fixed point to a certain contraction mapping. A measure of the inherent difficulty or computational complexity of the dynamic programming problem is

the dimension of the associated fixed point problem. The NFXP algorithm has been successfully programmed on an IBM-PC and applied to estimate a model of bus engine replacement where the fixed point dimension was at most 180 (Rust 1987). By comparison, the fixed point dimension for the retirement problem can be as large as several million.

This paper shows how to apply the NFXP algorithm to the retirement problem and demonstrates how to exploit the algebraic structure of the fixed point problem in order to rapidly compute high-dimensional fixed points on parallel vector processors like the Cray-2. With this technology one can formulate more realistic models of retirement behavior. Section 12.2 reviews some of the empirical issues that motivated the construction of the model. Section 12.3 develops the model, formulating the retirement decision process as a discrete control process. Section 12.4 presents computational results which show that fixed points as large as several million dimensions can be rapidly and accurately calculated on the Cray-2. Future work (see Rust 1989) will use the NFXP algorithm and data from the longitudinal Retirement History Survey (RHS) to actually estimate the unknown parameters of the model.

12.2 Empirical Motivation for the Dynamic Programming Model

The a priori structure of the dynamic programming model has been heavily influenced by my interpretation of the extensive empirical literature on retirement and consumption/savings behavior that has appeared over the last twenty years. This section summarizes some of the basic empirical and policy issues of the retirement process that I wanted the model to capture.

12.2.1 Accounting for Unplanned Events and the Sequential Nature of Decision-Making

Several existing models, such as Anderson, Burkhauser, and Quinn (1984) and Burtless and Moffitt (1984), studied retirement behavior in the context of a two-period model that divided time into a preretirement and postretirement phase. At some initial planning date before retirement, the worker is assumed to choose a fixed optimal retirement date and fixed preretirement and postretirement consumption levels. Anderson, Burkhauser, and Quinn used data from the RHS survey to find out how closely workers followed their initial retirement plans. In the initial 1969 wave of the survey, nonretired workers reported their planned retirement age. By tracing workers over the subsequent ten years they were able to compare the actual and planned retirement dates, and found that over 40 percent of the initial sample deviated from their initial retirement plans by over one year.

Clearly workers do not make single, once-and-for-all plans about consumption levels and retirement date. Rather, workers are constantly modifying their plans in light of new information. Anderson, Burkhauser, and Quinn found that unexpected changes in health, labor market conditions, and government policy (Social Security regulations, in particular) were the most important factors leading to revised retirement plans. This suggests a stochastic dynamic programming formulation where the solution takes the form of an optimal decision rule that specifies workers' optimal consumption and labor supply decisions as a function of their current information.

12.2.2 Accounting for Bequests

Many of the early studies of the impact of Social Security on private saving were based on the life-cycle consumption hypothesis of Modigliani and Brumberg (1954). Under the simple life-cycle model with no bequests: (1) consumption is predicted to remain constant or increase with age (depending on whether the interest rate is greater than or equal to the subjective discount rate), (2) workers are predicted to run down their accumulated wealth to zero by their (certain) date of death, and (3) intergenerational transfers like Social Security displace an equal amount of private savings (a greater amount if there is a net wealth transfer, due to the wealth effect on consumption). Initial work using cross-sectional data (Mincer 1979, Danziger et al. 1982, Kurz 1984, and Menchik and David 1983) provided evidence that contrary to the simple life-cycle model, age-wealth profiles are constant (or possibly increase) with age, and "the elderly not only do not dissave to finance their consumption during retirement, they spend less on consumption goods and services (save significantly more) than the nonelderly at all levels of income" (Danziger et al. 1982, p. 224). A study of consumption profiles using the RHS data by Hamermesh (1984) found that on average consumption exceeds earnings by 14 percent early in retirement, but that workers respond "by reducing consumption at a rate sufficient to generate positive changes in net financial worth within a few years after retirement" (p. 1). A study of estimated earnings and consumption paths by Kotlikoff and Summers (1981) indicated that intergenerational transfers account for the vast majority of the capital stock in the United States, with only a negligible fraction attributable to life-cycle savings. Direct observations of bequests from probate records (Menchik and David 1985) showed that bequests are a substantial fraction of lifetime earnings. Their results also demonstrated that bequests are a luxury good, with a "marginal propensity to bequeath" that is about six times higher in the top wealth quintile than in the lower four quintiles. As a whole, these studies provide a strong case for including bequests in a properly specified empirical model.

The policy implications of bequests were first pointed out by Barro (1974). Barro's "equivalence result" shows that under general conditions consumers can offset the effects of government tax policy (such as Social Security) by corresponding changes in private intergenerational transfers. In particular, the net wealth transfers to Social Security beneficiaries during the 1970s are predicted to be completely offset by increases in private saving for bequests.

Recent theoretical and empirical research, however, has questioned the importance of bequests as a determinant of consumption behavior during retirement. Davies (1981) showed that in a model with imperfect annuities markets and uncertain lifetimes, risk averse consumers can continue to accumulate wealth during retirement through a precautionary savings motive even though there is no bequest motive. Given that lifetimes are not certain, this creates the empirical problem of distinguishing between intended and accidental bequests. Recent panel data studies by Diamond and Hausman (1984b), Bernheim (1984), and Hurd (1986) found that the elderly do dissave after retirement. Hurd found that average real wealth in the RHS decreased by 27 percent over the ten-year period of the survey and concluded that "there is no bequest motive in the RHS, and, by extension, in the elderly population with the possible exception of the very wealthy. Bequests seem to be simply the result of mortality risk combined with a very weak market for private annuities" (p. 35). Menchik and David's (1985) study also casts doubt on the empirical relevance of Barro's equivalence result. Their regressions of bequests on gross Social Security wealth and the lifetime wealth increment, LWI (the difference between the discounted value of Social Security receipts and Social Security taxes), produced no evidence that bequests increase to offset increases in LWI; in fact, those in the top wealth quintile appeared to decrease bequests in response to an increase in LWI. However, their results also cast doubt on the Davies variant of the life-cycle model. To the extent that Social Security is a replacement for an incomplete annuities market, one would expect that gross Social Security benefits would decrease accumulated private wealth and unanticipated bequests. Menchik and David found a positive (albeit statistically insignificant) coefficient on gross Social Security benefits, and concluded that the "results indicate no significant effect of Social Security wealth on the age-wealth profile, a finding at odds with the life-cycle hypothesis. We find that Social Security does not depress or displace private saving and that people do not deplete their private assets in old age as is commonly assumed" (p. 432).

These conflicting theoretical and empirical results suggest the need to build a model that allows for both uncertain lifetimes and a bequest motive. A unified treatment may help to sort out their separate effects on the path of consumption during retirement. However, the fact that

bequests are not needed to explain the slow rate of wealth decumulation suggests that it will be very difficult to separately identify workers' subjective discount factors, the parameters of their bequest functions, and their subjective mortality probability distributions.

12.2.3 Accounting for the Joint Endogeneity of Labor Supply and Savings Decisions

The decline in the labor force participation rate of older males over the past thirty years is a well-known phenomenon; the participation rate for workers aged 55–64 declined from 86.8 percent in 1960 to 72.3 percent in 1980, and the rate for workers aged 65+ declined from 33.1 percent to 19.1 percent over the same period. Many people have blamed this decline on the historical increase in Social Security retirement benefits, which increased in real terms by more than 50 percent from 1968 to 1979, the decade of the RHS survey. Savings rates have also declined in the postwar era, from an average of 8.8 percent in the 1950s, 8.7 percent in the 1960s, 7.7 percent in the 1970s, to only 5.1 percent since 1980. Some researchers, including Feldstein, have claimed that Social Security “depresses personal saving by 30–50 percent” (Feldstein 1974, p. 905). However, according to economic theory an actuarially fair Social Security program should have no effect on aggregate savings or labor supply decisions; instead, simply inducing a 1-for-1 displacement of private savings by public savings (Crawford and Lilien 1982). It is well known, however, that the Social Security benefit formulas are not actuarially fair, but rather have strong incentives for early retirement (especially beyond age 65, see Burtless and Moffitt 1984). However, if workers increase their savings to prepare for earlier retirement, then the theoretical impact of Social Security on aggregate savings is ambiguous: the decreased savings due to the tax and wealth transfer effects may be offset by the increased savings due to the early retirement effect.

Empirical work designed to resolve these questions has failed to provide clear conclusions about Social Security's impact on labor supply and savings behavior. While analyses of labor supply decisions generally agree that Social Security does induce earlier retirement, there is substantial disagreement over the magnitude of the effect. Some studies such as Boskin and Hurd (1974) find a substantial impact, while others such as Sueyoshi (1986) find a moderate impact, and still others such as Burtless and Moffitt (1984) and Fields and Mitchell (1985) find a very small impact. In fact, the latter study found that a 10 percent decrease in benefits would increase the average retirement age by at most 1.7 months. Studies of Social Security's impact on aggregate savings are in disagreement about even the sign of the effect. For example, Barro (1978) used the same time series data as Feldstein (1974)

and an alternative measure of Social Security wealth, and found that increases in Social Security increased aggregate savings. He concluded that “the time-series evidence for the United States does not support the hypothesis that Social Security depresses private saving” (p. 1). Studies using longitudinal data such as Kotlikoff (1979) have generally found that Social Security reduces private saving, but have not found the 1-for-1 displacement of private savings that the simple life-cycle model predicts. Kotlikoff’s results show a partial offset ranging from 40 to 60 cents for every additional dollar of Social Security benefits; the increased savings due to early retirement did not turn out to be large enough to offset Social Security’s negative tax and wealth transfer effects.

A careful analysis of the impact of changes in Social Security benefits requires a model that treats labor supply and consumption as jointly endogenous decisions. Although a model that focuses on the last stage of the life-cycle probably will not be able to shed much light on Social Security’s impact on aggregate savings, it should address the historical decline in labor force participation of older men. The discrepancies in previous empirical results emphasize the need to carefully model the actuarial and benefit structure of the Social Security system, and if possible, to model workers’ expectations and uncertainties about changes in future benefits.

12.2.4 Accounting for Health and the Impact of Social Security Disability Insurance

Health problems are a major source of uncertainty in retirement planning, especially in terms of lost earning potential and unanticipated health care costs. Data from the NLS and RHS surveys indicate that poor health is a major factor in retirement decisions, especially among early retirees. Of the people retired in the 1969 wave of the RHS survey, 65 percent reported they were retired due to poor health; for those who had been out of the labor force for more than six years (the early retirees) the figure was 82 percent. Health problems are prevalent even among those who work; 39 percent of the 1969 RHS sample reported a health problem that limited their ability to work or get around, even though 63 percent of this group continued to work at a full- or part-time job. However, the inherent subjectivity of self-reported health measures and the financial incentives for claiming poor health in order to receive disability payments have led some to question the accuracy of health variables and the importance of poor health as a cause of retirement (Parsons 1982). In fact, some researchers (Bound 1986) have presented evidence (see figure 12.1) that suggests that much of the decline in the labor force participation rates of older males over the last thirty years can be ascribed to increases in disability claims allowed

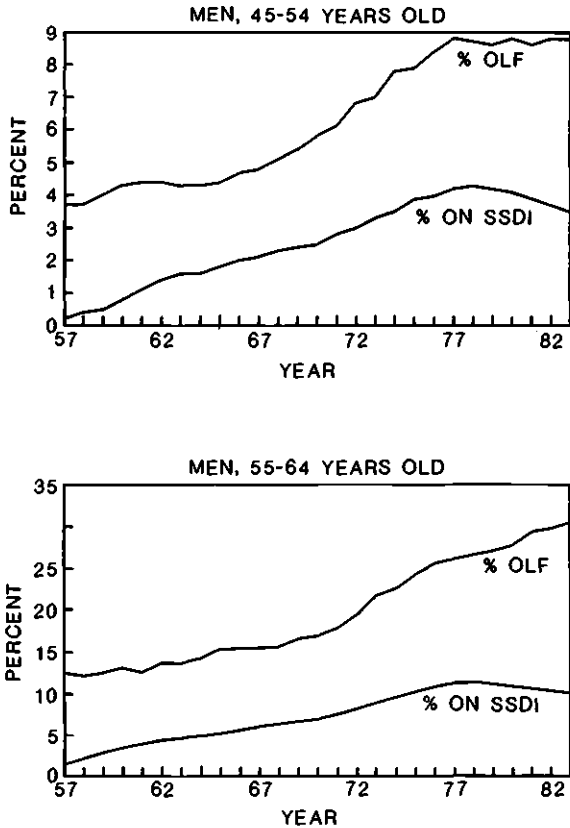


Figure 12.1

under the Social Security disability insurance program instituted in the late 1950s and substantially liberalized during the 1970s. Other researchers, such as Kotlikoff (1986), suggest that disability insurance may also be partly responsible for the decline in saving rates since it eliminates the need for precautionary saving to insure against unexpected illness or disability.

To the extent that qualification for disability insurance requires medical examination, the classification “disabled” is relatively more objective than self-reported measures of poor health. However, other approaches that use more “objective” measures of health status such as impairment indices (Chirikos and Nestel 1981), or ex post mortality (Parsons 1982, Mott and Haurin 1981), generally obtain results that are in broad agreement with studies that use self-reported measures of health status (although there are certain questions for which the alternative measures lead to important differences, see Chirikos and Nestel

1981, p. 113). Regardless of how it is measured, health status clearly has a significant impact on the labor force participation decision and appears to be one of the most important variables driving the dynamics of the retirement process. It is important, however, to find a measure of health status that does not rely heavily on subjective self-assessments, for example, classifying as disabled only those who have had doctor certification of disability (as is required in order to obtain disability benefits). The model must also incorporate the regulations and uncertainties governing the receipt of Social Security disability insurance; only by doing so can we hope to sort out the relative impact of liberalized disability vs. retirement benefits on the declining labor force participation rate of older males.

12.2.5 Accounting for "Partial Retirement" and Multiple Labor Force Transitions

Many models treat retirement as a dichotomous choice between full-time work and zero hours of work. However, economic theory suggests that workers might be better off if they could make a gradual transition from full-time work into retirement. Thus, at the other extreme are the labor supply models of Gordon and Blinder (1980) and MaCurdy (1983) that treat hours of work as a continuous choice variable. Gustman and Steinmeier (1983, 1984) have shown that a majority of non-self-employed workers face implicit or explicit minimum hours constraints that prevent them from gradually phasing out of their full-time jobs. Their analysis of the RHS data showed that approximately one third of all workers attempt to circumvent the minimum hours constraint through a spell of "partial retirement" in a part-time job. This suggests that a trichotomous choice model with the alternatives of full-time work, part-time work, and retirement, may be a better approximation to the actual choice sets facing workers than either the binary or continuous-choice formulations.

The RHS data show substantial variation in the paths workers follow into retirement. Table 12.1 presents the sequence of self-reported labor market states in the first four waves of the RHS. The table indicates that one needs at least a three-alternative choice set to adequately explain the variety of labor force transitions that occur along the path to retirement. It also indicates that the transition into retirement seems to be nearly an absorbing state; very few people "unretire" by re-entering a full-time job once fully or partially retired (or part-time job once fully retired). These numbers differ significantly from labor market re-entry rates presented by Diamond and Hausman (1984b) using NLS data. Table 12.2 reproduces their estimates of the fraction of men in the NLS survey that re-enter full-time work from the state of retirement or partial retirement. A possible explanation for the discrepancy is that

Table 12.1 Distributions of Retirement Sequences

Sequence	Frequency	Sequence	Frequency
ffrr	16.2%	frxx	1.6%
ffff	14.4	rrrx	1.5
fffr	11.2	fprr	1.4
rrrr	8.6	frxx	1.1
frrr	7.3	prrr	1.1
ffxx	5.4	ffrp	1.1
fffp	4.8	ffpx	0.7
ffpp	2.8	ffpf	0.6
ffpr	2.8	fprr	0.5
ffrx	2.5	pqrr	0.5
rrxx	2.2	others	9.8
fpr	2.1		

Source: Gustman and Steinmeier (1986, p. 566). The first letter in the retirement sequence is the individual's status in 1969, the first year of the RHS. The second, third, and fourth letters indicate their status in 1971, 1973, and 1975, respectively. The notation of the letters is: f = working full-time, p = working part-time, r = fully retired, x = status indeterminate. Sequences with a frequency less than 0.5 percent were grouped in the category "others".

Table 12.2 Labor Market Re-entry Rates

Age	One-Year Re-entry Rates		Two-Year Re-entry Rates	
	Self-described Retired or Unable to Work	Not Full-time Worker	Self-described Retired or Unable to Work	Not Full-time Worker
45-59	18.54	52.55	4.00	53.76
50-54	16.23	46.93	17.68	41.03
55-59	15.94	31.85	10.31	25.23
60-64	13.37	15.45	9.57	7.15
65-69	11.74	5.02	9.04	2.94
Total	14.53	29.48	10.13	16.72

Gustman and Steinmeier used a self-reported measure of labor force status to construct table 12.1.² The concept of retirement is ambiguous: Is someone who quits their full-time career job and takes a part-time job retired? Workers may interpret the concept differently and respond differently even though they are in identical labor force states. This suggests the use of objective measures of labor force status based on reported hours of work. Furthermore, from a modelling standpoint it seems undesirable to impose a priori constraints such as making retirement an absorbing state or prohibiting various transitions to and from different labor market states. The model should have the flexibility

to allow the data and the estimated parameter values “explain” what types of transitions actually occur.

Developing a tractable empirical model that incorporates all of these features is a challenging undertaking. Certainly a unified model will lack some of the fine detail of previous models that focused on specific aspects of the retirement process. However, the most important cost is the computer time required to solve and estimate the model. To my knowledge there is no simple analytic solution to the model I present in the next section: it seems to require numerical solution, a substantial computational task. Before presenting the model, I should answer a natural question: Isn't there a better way to estimate the model than by “brute force” numerical solution of the dynamic programming problem? In particular, MaCurdy (1983) developed a relatively simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty. Why not use MaCurdy's method? MaCurdy's approach is not well-suited to the retirement problem due to his assumption that consumption and hours of work are continuous choice variables. This allows MaCurdy to derive first-order conditions for the stochastic dynamic programming problem that equate the marginal rate of substitution between consumption and leisure to the real wage rate. This provides a computationally convenient orthogonality condition to estimate the identified parameters of the model. Unfortunately, the method depends critically on the assumption that workers do not face minimum hours constraints in their full-time jobs, and that one always has an interior solution with positive values for consumption and hours of work. MaCurdy recognizes this: “Because the procedure ignores statistical problems relating to the endogeneity of labor decisions, [it is] of limited use in estimating period-specific utilities associated with households in which corner solutions for hours of work are not a certainty . . . such as households with wives and older households where retirement may occur” (MaCurdy 1983, p. 277). The next section presents a model and estimation algorithm that can accommodate minimum hours constraints and corner solutions, but at the cost of repeated numerical solution of the dynamic programming problem over the course of the maximum likelihood estimation procedure.

12.3 Formulation of the Dynamic Programming Model

This section presents a theoretical model of retirement behavior that attempts to account for some of the empirical issues raised in section 12.2. The ultimate goal is to estimate and test the model using the RHS panel data. The primary factors limiting the realism of the model are computational feasibility and the availability of good data. The

construction of the theoretical model reflects these practical constraints. In particular, the RHS has limited data on private pension plans, so I restrict the model to male heads of household with no private pensions. Given the negligible use of private annuities and health plans among RHS respondents, it follows that Social Security is the predominant source of both retirement and health insurance benefits for this subsample.

12.3.1 State and Control Variables

In order to capture the fundamental dynamics of retirement behavior the model should include the following state variables which directly or indirectly affect workers' realized utility levels:

- w_t : accumulated financial and nonfinancial wealth
- y_t : total income from earnings and assets
- aw_t : the Social Security average monthly wage
- 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)

The state variables represent a subset of workers' current information that affects their expectations about their remaining lifespan, future earnings and retirement benefits, and their future health and family status. Since Social Security retirement and disability benefits are determined from the worker's primary insurance amount (a function of aw_t , which is in turn a complicated weighted average of past earnings), the variable aw_t summarizes the worker's expectations of future benefits accruing to him in retirement or disability, assuming fixed Social Security rules governing timing and eligibility for benefits. Since it is very difficult to formulate a low-dimensional state variable representing how the Social Security benefit structure changes over time, I assume that workers had "semi-rational" expectations of the benefit structure, equal to the regulations in force as of 1973. Although real benefits increased 51.2 percent between 1968 and 1979, the majority of the increase, 46.7 percent, was in effect by 1973 (see Anderson, Burkhauser, and Quinn 1984). The 1973 Social Security Act also changed the "earnings test" to reduce the 100 percent tax on earnings beyond the previous earnings limit to a 50 percent tax on all earnings over \$2,100. I describe the expectations assumption as semi-rational because I assume that workers correctly anticipated the cumulative changes in Social Security that came into effect over the period 1969-73, but maintained static expectations that no further changes would occur thereafter.

Given their expectations, at each time t workers must choose values of the following control variables:

- d_t : the employment decision (full-time/part-time/exit labor force)
- c_t : the level of planned consumption expenditures

The workers' sequential decision problem is to choose at each time t values for the control variables $i_t \equiv (c_t, d_t)$ that maximize the expected discounted value of utility over their remaining lifetime, where expectations are conditioned by the current values of the state variables $x_t \equiv (w_t, h_t, a_t, ms_t, e_t, y_t, aw_t)$. The goal is to specify a model that is parsimonious, yet rich enough to allow for certain kinds of heterogeneity. Perhaps the most important source of heterogeneity is differences in workers' attitudes toward retirement. Some workers may be "workaholics" who prefer working to the idle leisure of retirement, whereas others are "leisure lovers" who would jump at the chance to quit their jobs.

Notice that the formulation distinguishes between the worker's employment state and his employment decision. This feature allows the model to account for various labor force transitions, including "unretirement" and job search behavior, summarized in table 12.3.

12.3.2 Formulating Retirement Behavior as a Discrete Control Process

I model retirement behavior as a *discrete control process*, a discrete-time Markovian decision problem where the control variable is restricted to a finite set of alternatives. This framework represents workers' preferences as a discounted sum of a state-dependent utility function $u(x_t, i_t)$, and their expectations as a Markov transition probability $\pi(x_{t+1} | x_t, i_t)$. Blackwell's Theorem (Blackwell 1965, theorem 6) establishes that under very general conditions, the solution to a Markovian decision problem takes the form of a decision rule $i_t = f_t(x_t)$ that specifies the agent's optimal action i_t in state x_t . Note, however, that if the econometrician is assumed to observe the complete state vector x_t , this framework implies that knowledge of the true utility function u would enable him to solve for f and perfectly predict the agent's choice in each state x , producing a degenerate statistical model.³ A possible solution is to add an error term in order to obtain a nondegenerate statistical model of the form $i_t = f_t(x_t) + \eta_t$. Unfortunately, such ad hoc solutions are internally inconsistent: the economic model assumes that the agent behaves optimally, yet the statistical implementation of the model assumes that the agent randomly departs from optimal behavior. One wants a framework that can account for the fact that the agent has information ε_t that the econometrician does not observe. By

Table 12.3 Accounting for Labor Force Transitions in the Dynamic Programming Model

Employment state, e_t	Employment decision, d_t	Interpretation
ft	ft	Continue working at current full-time job
ft	pt	Quit current full-time job, search for a new part-time job
ft	ne	If $a_t \geq 62$, retire; if $a_t < 62$ and disabled, receive disability insurance; otherwise exit labor force
pt	ft	Quit current part-time job and search for a full-time job
pt	pt	Continue working at current part-time job
pt	ne	If $a_t \geq 62$, retire; if $a_t < 62$ and disabled, collect disability insurance; otherwise exit labor force
ne	ft	Unemployed, disabled, or retired worker searching for full-time job
ne	pt	Unemployed, disabled, or retired worker searching for part-time job
ne	ne	If $a_t \geq 62$, remain retired; if $a_t < 62$ and disabled, collect disability insurance; otherwise remain out of labor force

incorporating such unobserved state variables one obtains a nondegenerate, internally consistent statistical model generated by optimal decision rules of the form $i_t = f_t(x_t, \varepsilon_t)$. Rust (1988) developed a formal statistical framework for structural estimation of discrete Markovian decision problems with unobserved state variables. Table 12.4 summarizes the basic structure of the problem.

The solution to the decision problem consists of a sequence of decision rules or *controls* $f_t(x_t, \varepsilon_t)$ that maximize expected discounted utility over an infinite horizon. Define the value function V by

$$(1) \quad V(x_t, \varepsilon_t) = \sup_{\Pi} E \left\{ \sum_{j=t}^{\infty} \beta^{j-t} [u(x_j, f_j) + \varepsilon_j(f_j)] | x_t, \varepsilon_t \right\},$$

where $\Pi = \{f_t, f_{t+1}, f_{t+2}, \dots\}$, $f_t(x_t, \varepsilon_t) \in C(x_t)$ for all t , x_t , and ε_t , and where the expectation is taken with respect to the transition density for the controlled stochastic process $\{x_t, \varepsilon_t\}$ determined from Π and the transition density $p(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i)$. Under general conditions the value function V will be the unique solution to Bellman's equation

Table 12.4 Summary of Notation for Discrete Control Problem

Symbol	Interpretation
$C(x_t)$	Choice set: a finite set of feasible values for the control variable i , when the observed state variable is x_t
$\varepsilon_t = \{\varepsilon_t(i) i \in C(x_t)\}$	A $\#C(x_t)$ -dimensional vector of state variables observed by the agent but not by the econometrician; $\varepsilon_t(i)$ is interpreted as an unobserved component of utility of alternative i in time period t
$x_t = \{x_t(1), \dots, x_t(M)\}$	An M -dimensional vector of state variables observed by the agent and econometrician
$u(x_t, i) + \varepsilon_t(i)$	Realized single period utility obtained in state (x_t, ε_t) when alternative i is chosen
$p(x_{t+1}, \varepsilon_{t+1} x_t, \varepsilon_t, i)$	Markov transition density for next period state variable when alternative i is chosen and when the current state is (x_t, ε_t)

$$(2) \quad V(x, \varepsilon) = \max_{i \in C(x)} [u(x, i) + \varepsilon(i) + \beta EV(x, \varepsilon, i)],$$

where the function $EV(x, \varepsilon, i)$ is defined by

$$(3) \quad EV(x, \varepsilon, i) \equiv \int_y \int_{\eta} V(y, \eta) p(dy, d\eta | x, \varepsilon, i).$$

Blackwell's Theorem implies that the solution Π is stationary, $\Pi = \{f, f, f, \dots\}$, and Markovian so the agent's optimal decision rule $i = f(x, \varepsilon)$ depends only on the current values of the state variables determined by finding the alternative i that attains the maximum in Bellman's equation

$$(4) \quad f(x, \varepsilon) = \operatorname{argmax}_{i \in C(x)} [u(x, i) + \varepsilon(i) + \beta EV(x, \varepsilon, i)].$$

The sample likelihood function is derived from the conditional choice probabilities $P(i|x)$, which are obtained from the agent's optimal decision rule $i = f(x, \varepsilon)$ by integrating out over the unobserved state variable ε using the conditional density of ε given x . From equation (4) one can see that the unobservables enter nonlinearly in the conditional expectation of the value function, $EV(x, \varepsilon, i)$. Under standard distributional assumptions for the unobservables, ε_t will be continuously distributed on R^N , where $N = \#C(x_t)$. This raises serious computational difficulties, since calculation of $P(i|x)$ will ordinarily require N -dimensional numerical integration over ε in the optimal decision rule defined by (4). However, the expected value function $EV(x, \varepsilon, i)$ entering (4) will almost never have a convenient analytic formula, but must be computed by numerically integrating the value function V in (3). The value function must in turn be numerically computed by solving V as a functional fixed point to Bellman's equation (2). Since ε is a vector

of continuous state variables, it must be discretized in order to compute V on a digital computer. The discretization procedure approximates the true function V , an element of an infinite-dimensional Banach space B , by a suitable vector in a high-dimensional Euclidean space. Even with a very coarse grid approximation to the true continuous distribution of ε_t , the dimensionality of the resulting discrete approximation will generally be too large to be computationally tractable. These computational problems motivate the following assumption on the joint transition density for $\{x_t, \varepsilon_t\}$:

Conditional Independence Assumption: The Markov transition density factors as

$$(5) \quad p(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i) = q(\varepsilon_{t+1} | x_{t+1}) \pi(x_{t+1} | x_t, i), \quad i \in C(x_t).$$

This assumption involves two restrictions. First, x_{t+1} is a sufficient statistic for ε_{t+1} , which implies that any statistical dependence between ε_t and ε_{t+1} is transmitted entirely through the vector x_{t+1} . Second, the probability density for x_{t+1} depends only on x_t and i , and not on ε_t . Although (5) is a strong assumption, Rust (1988) developed a simple Lagrange multiplier statistic to test its validity. The payoff to assumption (5) is given by the following theorems of Rust (1988).

Theorem 1: Let $G(v(x)|x)$ denote the Social Surplus function, defined by

$$(6) \quad G(v(x)|x) \equiv \int \max_{i \in C(x)} [v(x, i) + \varepsilon(i)] q(d\varepsilon | x),$$

and let $G_i(v(x)|x)$ denote the partial derivative of $G(v(x)|x)$ with respect to $v(x, i)$. Then under the Conditional Independence assumption the conditional choice probability $P(i|x)$ is given by

$$(7) \quad P(i|x) = G_i(v(x)|x) \quad i \in C(x),$$

where the function v is the unique fixed point to the contraction mapping $v = T(v)$ defined by

$$(8) \quad T(v)(x; i) = u(x, i) + \beta \int_y G(v(y)|y) \pi(dy | x, i) \quad i \in C(x).$$

The function v is related to the value function V defined in (1) and (2) by

$$(9) \quad V(x, \varepsilon) = \max_{i \in C(x)} [v(x, i) + \varepsilon(i)].$$

Theorem 2: Under assumption (5) the controlled stochastic process $\{i_t, x_t\}$ is Markovian with transition density given by

$$(10) \quad Pr\{i_{t+1}, x_{t+1} | i_t, x_t\} = P(i_{t+1} | x_{t+1}) \pi(x_{t+1} | x_t, i_t).$$

Products of the transition density given in (10) form the likelihood function for the process $\{i_t, x_t\}$. This function is difficult to evaluate

primarily because the conditional choice probability $P(i|x)$ requires calculation of the value function v as a fixed point of the contraction mapping (8). Theorem 1 shows that from the standpoint of evaluating (10), there are two major payoffs to assumption (5). First, it implies that ε does not enter the expected value function $EV(x, \varepsilon, i)$, so that ε enters V only additively as shown in (9). This implies that the conditional choice probabilities $P(i|x)$ for the dynamic discrete choice model are given by exactly the same formulas as for static discrete choice models, except that the relevant utility function is not the static utility function u , but the fixed point v of the contraction mapping (8). Second, assumption (5) implies that the dynamic programming problem can be solved by computing the fixed point $v = T(v)$ over the space $\Gamma = \{(x, i) | x \in R^M, i \in C(x)\}$. This is a much easier task than computing the fixed point $V(x, \varepsilon)$ over the direct state space $\nabla = \{(x, \varepsilon) | x \in R^M, \varepsilon \in R^N\}$ since ε is a continuous-valued N -dimensional vector which must be discretized into K^N values (where K is the diameter of the grid for ε), whereas the argument i entering $v(x, i)$ is already discrete and assumes at most N values.

Given a parametric specification for the unknown objects u , q and π , one can "recover" the agent's underlying preferences (β, u) and expectations (π, q) by finding parameter values that maximize the likelihood function. This suggests the following *nested fixed point algorithm*: an "outer" nonlinear optimization algorithm searches for the parameter vector θ that maximizes the likelihood function, and an "inner" fixed point algorithm recalculates the fixed point v_θ of (8) each time the outer optimization algorithm updates its estimate of θ . Rust (1988) showed that under certain regularity conditions, the NFXP algorithm produces consistent and asymptotically normally distributed parameter estimates.

Before presenting parametric specifications for u , q and π , I should mention some drawbacks of the discrete control formulation. Although I have argued that there are good reasons for treating the employment decision d_t as discrete, both time t and the consumption decision c_t appear to be better approximated by continuous variables. My response is that the discrete formulation seems to be the best available approximation given the computational and data limitations I face. The computational limitation is that, to my knowledge, there are no estimation algorithms available for continuous-time stochastic control problems, or for dynamic programming models where the decision variable is mixed discrete/continuous.⁴ The data limitation is that individuals in the RHS are sampled at two-year intervals with only limited retrospective information on their states and decisions between survey dates. In theory, one could formulate a very fine grain, discrete-time model (regarded as a close approximation to the actual continuous-time decision process) and "integrate out" the dates for which no data are

available, but the computational burden required to solve the model and perform the integrations appears to be prohibitive. Therefore I interpret the decisions $i_t = (d_t, c_t)$ as “plans”, as of date t , that are revised at the same two-year time intervals as the survey dates. Thus, the state variables x_t refer to the worker’s state in the previous two years, and the decisions $i_t = (d_t, c_t)$ refer to the worker’s plans regarding consumption and labor force participation over the next two years. The plans need not be fulfilled, hence there will be a conditional probability distribution for the state x_{t+1} at time $t+1$ conditional on the current state x_t , and plan $i_t = (d_t, c_t)$ chosen at time t . Under this interpretation it is much more natural to regard the choice of a consumption plan c_t as an interval rather than a specific number since there will be unforeseen future events that cause actual consumption to deviate from the plan. The use of consumption intervals may also help mitigate the effects of the inevitable errors in variables in the constructed consumption data.⁵ Since I do not actually observe the consumption plan chosen by the worker in the RHS, in the empirical implementation of the model I will assume that the ex post realized consumption interval coincides with the ex ante plan.

12.3.3 Specification of Workers’ Preferences

Table 12.5 summarizes the formulation of the retirement problem as a discrete control process. Death, quite naturally, is treated as an absorbing state and the bequest function specifies the utility of entering this state. The dynamic programming problem proceeds by backward induction from the (uncertain) age of death over two-year intervals back to an initial age, 58, the age of the youngest respondent in the first wave of the RHS.

It remains to specify the functional forms for b , u , π , and q .⁶ The NFXP algorithm places no restrictions on the functional forms for b , u , and π , but computational tractability appears to require that the distribution of unobservables q be a member of McFadden’s (1981)

Table 12.5 Summary of the Retirement Decision Problem

Item	Notation
Choice set	$C(x) = \{1,2,3\} \otimes \{c^1, \dots, c^J\}$, $1 = \text{ft}$, $2 = \text{pt}$, $3 = \text{ne}$
Control vector	$i_t = (d_t, c_t)$; $d_t \in \{1,2,3\}$, $c_t \in \{c^1, \dots, c^J\}$
State vector (observed)	$x_t = (w_t, h_t, a_t, ms_t, cs_t, e_t, y_t, aw_t)$
State vector (unobserved)	$\varepsilon_t \equiv \{\varepsilon_t(i) i \in C(x_t)\}$, $\varepsilon_t(i) = \varepsilon_t(d, c)$
Bequest function	$b(w_t, ms_t, cs_t, \theta_1)$
Utility function	$u(d_t, c_t, e_t, h_t, a_t, ms_t, \theta_2)$
Transition density (x_t)	$\pi(x_{t+1} x_t, i_t, \theta_3)$
Transition density (ε_t)	$q(\varepsilon_t x_t, \theta_4) \sim \text{GEV}(C(x_t), \theta_4)$, see (11)
Parameter vector	$\theta = (\beta, \theta_1, \theta_2, \theta_3, \theta_4)$, $1 \times (1 + K_1 + K_2 + K_3 + K_4)$

generalized extreme value (GEV) family.⁷ The GEV family is closed under the operation of maximization, leading to convenient closed-form expressions for the social surplus function (6) and its derivatives, the choice probabilities (7). This feature greatly simplifies the NFXP algorithm, avoiding the numerical integrations that are normally required for other multivariate distributions. I chose a particular member of this family whose cumulative distribution function $Q(\varepsilon, \theta_4)$ is given below.

$$(11) \quad Q(\varepsilon, \theta_4) = \exp \left\{ - \sum_{\delta=1}^3 \left[\sum_{j=1}^J \exp \{ \varepsilon(\delta, j) / \theta_{4\delta} \} \right]^{\theta_{4\delta}} \right\},$$

$$0 \leq \theta_{4\delta} \leq 1, \quad \delta = 1, 2, 3.$$

Since the corresponding density q does not depend on x , it follows that the unobserved state variables are serially independent in this specification. Formula (11) includes the standard multivariate extreme value distribution as a special case when $\theta_{4\delta} = 1, \delta = 1, 2, 3$. The latter distribution satisfies the well-known IIA property: the components $\varepsilon(\delta, j)$ and $\varepsilon(d, c)$ are contemporaneously independent when $(d, c) \neq (\delta, j)$. When $\theta_{4\delta}$ are not all equal to 1 one obtains a pattern of contemporaneous correlation in the components of ε_t , represented by the choice tree shown in figure 12.2.

Thus, (11) allows correlation in the unobserved state variables affecting the consumption decision c_t , given the labor supply decision d_t , but assumes independence in unobserved state variables corresponding to different labor supply choices. Formula (11) yields the following nested logit formulas for the conditional choice probabilities:

$$(12) \quad P(d, c | x, \theta) = P(c | x, d, \theta) P(d | x, \theta),$$

where $P(c | x, d, \theta)$ and $P(d | x, \theta)$ are given by

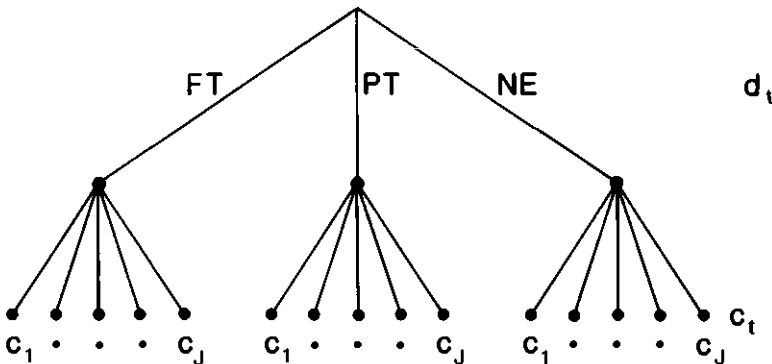


Figure 12.2

$$(13) \quad P(c|x, d, \theta) = \frac{\exp \{v_{\theta}(x, c, d)/\theta_{4d}\}}{\sum_{j=1}^J \exp \{v_{\theta}(x, j, d)/\theta_{4d}\}},$$

$$(14) \quad P(d|x, \theta) = \frac{\exp \{I(d)\theta_{4\delta}\}}{\sum_{\delta=1}^3 \exp \{I(\delta)\theta_{4\delta}\}},$$

where the *Inclusive Value*, $I(d)$, is defined by

$$(15) \quad I(d) = \ln \left[\sum_{j=1}^J \exp \{v_{\theta}(x, d, j)/\theta_{4d}\} \right] \quad d = 1, 2, 3.$$

Finally (11) yields an explicit formula for the fixed point condition (8)

$$(16) \quad v_{\theta}(x, d, c) = u(x, d, c, \theta_2) + \beta \int_y \ln \left[\sum_{\delta=1}^3 \left[\sum_{j=1}^J \exp \{v_{\theta}(y, \delta, j)/\theta_{4\delta}\} \right]^{\theta_{4\delta}} \right] \pi(dy|x, d, c, \theta_3),$$

with the implicit "terminal condition" that $v_{\theta}(x, d, c) = b(w, ms, \theta_1)$ if $h = \text{"dead"}$.

It remains to specify the functional forms for the bequest and utility functions, b and u . I assume that the bequest function has the following functional form

$$(17) \quad b(w, ms, \theta_1) = w^{\theta_{11}}(\theta_{12} + \theta_{13}ms).$$

The coefficient θ_{11} will reflect a diminishing or increasing marginal utility of bequests depending on whether θ_{11} is greater or less than 1. The Menchik and David study discussed in section 12.2.2 suggests that possibly $\theta_{11} > 1$. Presumably a married worker obtains greater utility from bequests to the remaining spouse than from bequests to friends, institutions, or the government. Thus, I expect that θ_{13} is positive.

The utility function is slightly more complicated. I assume that it has the following functional form:

$$(18) \quad u(d, c, e, h, a, ms, \theta_2) = \left[\sum_{i=1}^3 \sum_{j=1}^3 \theta_{2,ij} I\{d = i, e = j\} \right] [c^{\theta_{210}}] [e^{\theta_{211}}] [\theta_{212} + \theta_{213}a + \theta_{214}ms + \theta_{215}h].$$

According to (18), utility is a function of consumption c and the level of leisure e . Ranking the employment states as 1 = full-time (*ft*), 2 = part-time (*pt*), and 3 = not employed (*ne*), I expect the coefficient θ_{211} should be negative for a leisure lover and positive for a "workaholic".

The coefficient θ_{210} should be positive and less than 1 if there is diminishing marginal utility of consumption. The basic utility obtained from consumption and leisure is modified by the last factor in (18) which accounts for health status, age, marital status, and the presence of children. Ranking the health states as 1 = good health, 2 = fair health, 3 = disabled, I expect that the coefficient θ_{215} should be negative; being in worse health diminishes the utility obtained from consumption or leisure (or work, if he is a work-lover). It is not clear what sign to expect for the coefficient θ_{213} on the age variable. Perhaps as one gets older, one's remaining lifetime becomes more precious, suggesting a positive coefficient. However, aging might also result in general mental and physical deterioration independent of that captured by the health variable, suggesting a negative sign. One would ordinarily expect the presence of a spouse to increase the worker's utility, suggesting a positive value for θ_{214} .

The final term in (18) is the double summation term that reflects the monetary and psychic search costs of changing employment states. Perhaps the hardest transition to make is from the retired state to finding a new full-time job. This suggests the coefficient on $I\{d=1, e=3\}$ should be a large negative number, reflecting the data in table 12.1 that very few retired workers ever "unretire" and return to work at a full-time job. On the other hand, it should be relatively easy to make the reverse transition and retire from either a full- or part-time job: $I\{d=3, e=1\}$ or $I\{d=3, e=2\}$. Thus, the coefficients on these terms should be positive, possibly reflecting the utility value of any retirement bonuses or incentives. I would also expect that it is relatively easier to move into a part-time job from a full-time job, than vice versa, so I expect the coefficient on $I\{d=2, e=1\}$ to exceed the coefficient for $I\{d=1, e=2\}$. To the extent that workers desire to make a gradual transition from work to retirement, the coefficient on $I\{d=2, e=1\}$ should be positive, reflecting the prevalence of partial retirement discussed in section 12.2.5. The remaining coefficients reflect the utility costs of decisions to remain in the current employment state: $I\{d=3, e=3\}$, $I\{d=2, e=2\}$, and $I\{d=1, e=1\}$. For leisure lovers, there should be disutility associated with the decision to continue working, hence I expect the coefficient on $I\{d=1, e=1\}$ to be negative, but substantially less than coefficients for $I\{d=1, e=2\}$ or $I\{d=1, e=3\}$. The workers who partially retire might enjoy the experience, so it is possible that the coefficient for $I\{d=2, e=2\}$ is positive. In any case, it should be easier to remain on a current part-time job than to find a new one, so the coefficient for $I\{d=2, e=2\}$ should exceed the coefficients for $I\{d=2, e=1\}$ or $I\{d=2, e=3\}$. Of all the decisions, it is perhaps easiest to remain retired; thus, at least for leisure lovers, I expect that the coefficient for $I\{d=3, e=3\}$ to be positive.

The primary source of population heterogeneity that I wish to account for is the distinction between work lovers and leisure lovers. Rather than treat this as unobserved heterogeneity, one can use the responses from attitudinal questions in the RHS to classify each worker as a work lover or leisure lover, interacting this taste variable with the coefficients of u that can be expected to differ between work lovers and leisure lovers. One can account for additional heterogeneity by making certain parameters functions of time-invariant sociodemographic variables, the most important of which are race and the worker's main career occupation and industry.

12.3.4 Specification of Workers' Expectations⁸

Having specified the general form of the worker's per period objective function, it remains to specify the law of motion for the state variables. I assume that the observed state vector $x_t = (w_t, h_t, a_t, ms_t, e_t, y_t, aw_t)$ evolves according to a parametric Markov transition density $\pi(x_{t+1}|x_t, i_t, \theta_3)$ that depends on the worker's consumption and labor supply decision $i_t = (d_t, c_t)$. The transition density embodies the worker's expectations about his future health, his lifespan, and the future levels of his income and stock of wealth. More precisely, I assume that workers' individual expectations about future values of the state variables coincide with the population behavior of these variables (as represented by the estimated transition density π) within each sociodemographic stratum.

Since the transition probability π only depends on observable variables $\{i_t, x_t\}$, one could in principle use nonparametric methods to estimate it. With a discrete state space, the nonparametric estimate of $\pi(x_{t+1}|x_t, i_t)$ is simply the number of transitions (x_{t+1}, x_t, i_t) divided by the total number of transitions of the form (y, x_t, i_t) summed over all states y . However, with a large number of discrete cells and a limited amount of data, the nonparametric estimate of π will be identically zero for many transitions (x_{t+1}, x_t, i_t) , even though it is clear that such transitions can actually occur with positive probability. Therefore it is preferable to use parametric functional forms for π that "smooth out" the data on state transitions to yield positive estimates for all transition probabilities that are logically possible. It is also desirable to use flexible functional forms that do not impose arbitrary a priori restrictions on possible transitions. The conditional logit model (with full sets of alternative-specific dummies and sufficient terms for interactions of different explanatory variables) is an ideal candidate. However, given the very large number of possible states for x_{t+1} , a single joint estimation of π is out of the question. It is much simpler to decompose π as a product of conditional probabilities for each component $x_t(m)$, resulting in a series of tractable conditional logit estimations where the number of alternatives equals the (relatively small) number of values

that each component $x_i(m)$ can assume. Since a multivariate probability density can always be decomposed as a product of the conditional and marginal densities of its components, there is no loss in generality in this approach.

The state variable a_t , representing the worker's age has the simplest law of motion: $a_{t+1} = a_t + 2$. To keep a_t in a finite number of cells, I will assume that there is a maximum age of, say, 98 years which is treated as an absorbing state. This does not necessarily imply that all people die with probability 1 at age 98, rather the model simply does not account for further increases in the mortality hazard beyond age 98. For all practical purposes, however, the mortality rate for men over age 90 is so high that there is no effective loss in generality from assuming that all workers die with probability 1 at age 98, an assumption that leads to substantial computational simplifications, as I show in section 12.4. Therefore I assume that life ends with probability 1 at age 98 or before, implying that a_t takes on twenty values in increments of two from a starting age of 58.

The state variable h_t , representing health takes on one of four values, $\{1,2,3,4\}$, where 1 denotes good health, 2 denotes a health condition which the respondent reports limits his ability to get around or to work (yet which is not so severe that the worker is actually disabled), 3 denotes that the worker has been certified by a doctor to be disabled (and hence is not working and is eligible for Social Security disability benefits), and 4 denotes the absorbing state of death. States 1 and 2 are obviously somewhat subjective in nature. State 3, on the other hand, is much less subjective since Social Security has fairly strict rules regarding doctor certification of disability in order for a worker to receive disability benefits. According to Social Security rules, any person receiving disability benefits cannot work (except for a brief "trial employment" period lasting at most twelve months), so the employment state for a person with $h_t = 3$ should be the singleton $e_t = \{ne\}$. It is possible, however, for a disabled person to try to search for a job at the risk of losing his disability benefits. Thus, even though a person is disabled I allow the worker the full set of employment decisions, $d_t = \{ft, pt, ne\}$. This allows me (in at least a crude way) to study the effect of disability insurance on workers' incentives to re-enter the labor force.

Transitions between health states 1, 2, and 3 obey a parametric transition probability of the form

$$(19) \quad \pi_1(h_{t+1} | h_t, a_t, m, s_t, e_t, w_t, c_t, d_t; \theta_{31}),$$

which gives the probability of health next period as a function of health this period together with age, marital status, employment status, wealth, and the labor supply and consumption decisions. The function π can be taken to have a trinomial logit form, with separate coefficients for each

of the independent variables and their interactions. The estimated health transition probability can be interpreted as accounting for workers' perceptions of the "leniency" of admission to the disability program. For example, the conditional probability that $h_{t+1} = 3$ given $h_t = 2$ equals the worker's chances of getting onto the disability roles given that he is not in good health at time t . A separate binomial logit probability function captures workers' mortality assessments as a function of their age and other state variables x_t .

A binomial logit probability function will also be used to capture the stochastic process for worker marital status, ms_t , as a function of the state variables x_t and decision variables (c_t, d_t) . Marital status takes on two states, married or single. A married man may lose his wife through death or divorce, but once single, is allowed to remarry.

The state variable e_t representing the worker's employment status takes on three values $\{1, 2, 3\}$, corresponding to full-time work, part-time work, and not in the labor force, respectively. The conditional probability density for e_{t+1} has a trinomial logit form

$$(20) \quad \pi_2(e_{t+1} | e_t, a_t, ms_t, w_t, h_t, y_t, d_t; \theta_{32}).$$

It is particularly important to allow for the effects of age and health on re-employment probabilities. Wealth and income are included as proxies for unobserved job skills which may make the worker more employable; presumably wealthier, higher income workers have better job skills and are thus more employable. I include last period employment status e_t to control for any structural state dependence due to past lapses into unemployment or retirement. Presumably there is more stigma to being unemployed rather than retired, so an unemployed worker might face lower probabilities of re-employment than a retired person. Thus, the model might be able to provide some insight into the "discouraged worker effect" wherein a worker decides to retire rather than face the frustration of trying to search for a new job. I expect that a worker's chance of being fired from his current job increases with age and poor health, and decreases with "experience" as proxied by his current income and wealth, y_t and w_t . I also expect that full-time jobs are more secure than part-time jobs. Mandatory retirement beyond a certain age can be incorporated as a probability 1 chance of being fired when a_t exceeds the mandatory retirement age.

It remains to describe the transition probability function for wealth, w_t . The standard budget equation is that wealth next period equals wealth this period plus earnings and income from investments, less consumption expenditures:

$$(21) \quad w_{t+1} = w_t + y_t - c_t.$$

Thus, predicting next period's wealth reduces to predicting next period's earnings conditional on a specific choice of consumption interval

c_t . This requires estimating a transition density for total income y_t of the form

$$(22) \quad \pi_3(y_{t+1} | y_t, w_t, aw_t, h_t, a_t, e_t, ms_t, d_t; \theta_{33}).$$

Here the function π_3 can be thought of as an earnings function which predicts the worker's earnings and investment income over the next two years as a function of his observed state x_t (including his last period income y_t) and employment decision d_t . The earnings function captures workers' expectations about their future earnings streams and the retirement or disability benefits due to them under Social Security. For example, if the worker is currently employed full-time ($e_t = 1$), then π_3 will predict his next period earnings on his job. These earnings will be a function of his age, health, and level of job experience. Wealth w_t and income y_t are included as proxies for job skills, since presumably wealth, job earnings, and ability/experience are highly correlated. π_3 also includes investment income on existing wealth, $\bar{r}w_t$, where \bar{r} is a random rate of return on the worker's investment portfolio. Wealth will be measured to include both real and financial wealth, including real estate, the cash value of insurance policies, and other personal property such as automobiles and furniture, etc. If the worker is currently unemployed and searching for work, ($d_t = 2$ or $d_t = 1$, and $e_t = 3$), then the earnings function predicts the worker's unemployment insurance benefits. If the worker is retired, ($e_t = 3$ and $a_t \geq 62$), then π_3 predicts the worker's Social Security benefits. These benefits are a function of the worker's average monthly wage, aw_t (which determines his primary insurance amount and benefits), and his marital status, ms_t . π_3 also predicts payments from Social Security disability insurance and Medicare in the event the worker is disabled or in bad health, and the death benefit in the event the worker dies. Thus, the earnings function π_3 completely embodies the worker's expectations of his future earnings streams under all eventualities—retirement, employment, or unemployment—and includes contingent payments for health and life insurance. Changes in Social Security policy, such as changes in benefit levels or retirement ages, can be represented through appropriate changes in the earnings function. One can simulate the effects of changes in Social Security policy by appropriately altering the earnings function π_3 and recomputing the new optimal retirement strategy. This allows one to quantify how much workers are "hurt" by a policy change by measuring the lump-sum fee workers would be willing to pay in order to keep the existing Social Security rules intact. One can also measure how the policy change alters the probability of retirement for each configuration of the state variables.

The final state variable is the average monthly wage, aw_t . As an average of lifetime earnings, aw_t will be fairly insensitive to earnings levels and labor supply choices at the end of the worker's career,

especially once it is discretized. Thus, there are no real dynamics for aw_t ; it is simply an indicator of the level of benefits coming due to the worker. There is some question as to whether the average wage need even be included in the model since it should be very highly correlated with the earnings y_t on the worker's full-time job and wealth w_t . This is an empirical issue. If aw_t can be adequately proxied by y_t and w_t , I would eliminate it as a state variable to conserve on the dimensionality of the fixed point problem.

12.4 Numerical Computation of the Dynamic Programming Model

As described in section 12.3, the revealed preference problem reduces to estimation of the unknown parameter vector $\theta = (\beta, \theta_1, \theta_2, \theta_3, \theta_4)$, where $0 < \beta < 1$ is the worker's intertemporal discount factor, θ_1 are the parameters entering the bequest function, b , θ_2 are the parameters entering the utility function u , θ_3 are the parameters of the transition probability for the observed state variables π , and θ_4 are the parameters of the transition probability for the unobserved state variables q . The unknown parameters can be estimated by maximum likelihood method using the following three-step procedure.

Step 1. Estimate the vector θ_3 , entering the transition density $\pi(x_{t+1}|x_t, i_t, \theta_3)$ using the partial likelihood function $L_1(\theta_3)$ defined by

$$(23) \quad L_1(\theta_3) \equiv \prod_{k=1}^K \prod_{t=1}^5 \pi(x_{t+1,k}|x_{t,k}, i_{t,k}, \theta_3),$$

where k indexes individuals in the RHS sample.

Step 2. Using the initial consistent estimate of $\hat{\theta}_3$ from step 1, estimate $(\beta, \theta_1, \theta_2, \theta_4)$ using the partial likelihood function $L_2(\beta, \theta_1, \theta_2, \hat{\theta}_3, \theta_4)$ defined by

$$(24) \quad L_2(\beta, \theta_1, \theta_2, \hat{\theta}_3, \theta_4) \equiv \prod_{k=1}^K \prod_{t=1}^5 P(i_{t,k}|x_{t,k}, (\beta, \theta_1, \theta_2, \hat{\theta}_3, \theta_4)),$$

where P is defined by (12) through (15) and the fixed point condition (16).

Step 3. To get correct estimated standard errors and asymptotically efficient parameter estimates for $\hat{\theta}$, compute one Newton-step from the initial consistent estimate $\hat{\theta}$ using the full likelihood function $L_f(\theta)$ defined by

$$(25) \quad L_f(\theta) \equiv \prod_{k=1}^K \prod_{t=1}^5 P(i_{t+1,k}|x_{t+1,k}, \theta) \pi(x_{t+1,k}|x_{t,k}, i_{t,k}, \theta_3).$$

The nested fixed point algorithm is required only in steps 2 and 3 in order to compute the value function v_θ entering the conditional choice

probabilities P . This requires recomputing the fixed point v_θ of the contraction mapping (16) each time the outer nonlinear maximization algorithm computes new values for θ . As discussed in section 12.3, if there are continuous state variables, then the fixed point v_θ is an element of an infinite-dimensional Banach space B . The computational strategy is to discretize the continuous state variables, and in effect, approximate the infinite-dimensional space B by a high-dimensional Euclidean space R^N . The dimension of the fixed point problem N is equal to the number of possible values of i and the discretized values that x can assume. Suppose that w_i is discretized into 100 cells, y_i into 5 cells, and c_i into 5 cells. Assuming that aw_i can be proxied by w_i and y_i , the remaining state variables assume the following number of values: $h_i:4$, $a_i:20$, $e_i:3$, $ms_i:2$, $d_i:3$. The implied fixed point dimension is $N = 3,600,000 = 100*5*5*4*20*3*2*3$. Thus, a 3.6 million dimensional fixed point must be repeatedly recalculated in the fixed point subroutine of the nested fixed point algorithm during the course of the parameter search. It is therefore necessary to find algorithms to compute high-dimensional fixed points as rapidly as possible, say, in less than thirty seconds on a supercomputer such as the Cray-2.

By theorem 1 the fixed point problem can be written as $v = T(v)$, where the contraction operator T is defined in formula (16). There are two principal algorithms for computing contraction fixed points: contraction iterations and Newton-Kantorovich iterations. Contraction iterations involve repeated evaluations of the contraction mapping T starting from an arbitrary initial estimate v_0 :

$$(26) \quad v_{k+1} = T(v_k)$$

The Newton-Kantorovich method converts the fixed point problem into the problem of finding a zero of a nonlinear operator, $(I - T)(v) = 0$, where I is the identity operator on B and 0 is the zero element of B . This nonlinear equation is then solved for v using Newton's method:

$$(27) \quad v_{k+1} = v_k - [I - T'(v_k)]^{-1}[I - T](v_k)$$

where $T'(v_k)$ is the Fréchet derivative of T with respect to v evaluated at the point v_k . The method of successive approximations is guaranteed to converge for contraction mappings, however, the convergence is very slow (especially when β is close to 1). Newton's method has a very rapid quadratic rate of convergence, however the method is only guaranteed to work in a "domain of attraction" of points sufficiently close to the true fixed point v . The other disadvantage of Newton's method is that one must solve an $N \times N$ linear system involving the matrix $[I - T'(v_k)]$. For large N the time and storage required to solve the linear system becomes prohibitive.

Although the fixed point problem looks formidable at first glance, the retirement problem has a special structure that can be exploited in order to dramatically reduce the computational burden. There are two principal features of the retirement problem that can be effectively exploited: (1) using the absorbing state of death to induce a backward recursion for the value function, and (2) exploiting the sparsity structure of the transition probability matrix representation of π , in particular exploiting the deterministic transitions for a_t and the banded structure of the wealth transition probabilities. The first feature is based on the observation that in the absorbing state of death the value function has an a priori known functional form, $v_\theta = b$. Therefore v_θ need only be calculated for the three remaining health states, reducing the effective dimension of the problem from $N = 3.6$ million to $N = 2.7$ million. Under the additional assumption that workers die with probability 1 beyond some fixed age, say 98, one can compute the fixed point v_θ in a single contraction iteration, essentially by backwards induction from the last year of life (in this case, age 98). From an economic perspective, this is a relatively innocuous assumption since extremely few males live beyond age 98. However, without this assumption one is faced with an infinite-horizon problem since the model places no upper bound on the lifespan of the worker. In this case a combination of contraction and Newton-Kantorovich iterations are required in order to compute v_θ , increasing the required computer time by several orders of magnitude. Since the assumption of fixed lifespan is basically harmless and leads to substantial computational simplifications, I will adopt it in my empirical work. The second feature, exploiting the sparsity structure, allows one to economize on the number of storage locations required to hold the matrix representation of π , and to significantly reduce the number of operations needed to evaluate the contraction mapping T (16) or solve the $N \times N$ linear system in the Newton-Kantorovich iteration (27).

To understand the latter point, consider the work involved in computing a single evaluation of the contraction mapping T . Once the state vector x is discretized, the majority of the work is the required integration with respect to the transition probability π . This is equivalent to left matrix multiplication of the "vectorized" integrand by the matrix representation of π . Matrix multiplication is a very simple operation that is easily vectorized for maximum efficiency on a vector processor like the Cray-2. However, such matrix-vector multiplications require order N^2 multiplications and additions, where N is the number of discrete cells that x_t can assume. Even a machine that can multiply at 400 megaflops (400 million floating point operations per second) can get quickly bogged down when N exceeds several hundred thousand. It is therefore essential to reduce the total number of multiplications by

exploiting the sparsity of the matrix representation of π . Unfortunately, standard algebraic techniques for sparse matrices typically do not perform well on vector processors owing to the irregular memory reference patterns for their elements, creating "bank conflicts" that prevent the processors from running at maximum efficiency with continuously full vector pipelines. For example, even after extensive modification and optimization of standard sparse linear equation solvers, the resulting code typically runs slower than 12 megaflops on the Cray-1 (Duff 1984). This is significantly slower than the Cray-1's peak rates of 160 megaflops on dense linear algebra problems. The trick, then, is to exploit the sparsity structure of the transition matrix to reduce the total number of operations while at the same time attempting to keep the nonzero elements in a "locally dense" configuration so they can be fed to the vector registers in a continuous stream, allowing the processors to run uninterrupted at nearly peak speed.

Figures 12.3 through 12.8 depict different sparsity patterns for the matrix representation of π depending on the ordering of the component state variables in x_t . π can be regarded as a direct product of three types of transition matrices: (1) a circulant matrix for a_t , (2) a banded matrix for w_t , and (3) a dense matrix representing the joint transition matrix for the remaining state variables. These component matrices are depicted in figure 12.3. By varying the order of these component matrices in the construction of the direct product, one obtains different sparsity patterns for π . Figures 12.4, 12.5, and 12.6 depict the sparsity patterns for the orderings (d, w, a) , (w, a, d) , and (a, d, w) , respectively. None of these orderings is particularly desirable, for they all lead to fairly irregular and dispersed memory reference patterns. Figure 12.7 depicts the "optimal" sparsity pattern, (a, w, d) , which produces the maximum amount of local density in the storage pattern for the matrix elements. The matrix-vector multiplication under this structure occurs in an outer do-loop over age values 1 to 20, calling a block-banded matrix multiplication subroutine specially designed to keep the vector pipelines continuously full. Figure 12.8 shows the packed storage arrangement for the block-banded matrices that form the off-diagonal sectors of π . This arrangement allows one to fully exploit the sparsity of π while keeping the vector processor running at nearly maximum efficiency.

Exploitation of sparsity patterns is particularly important in the infinite horizon case. For sufficiently high discount factors, β , it will be optimal to use Newton-Kantorovich iterations rather than contraction iterations alone, but the former requires the solution of the linear system involving the matrix $[I - T'(v)]$. However it is easy to see that $T'(v)$ is simply β times the transition probability matrix for the controlled process $\{i_t, x_t\}$ which is isomorphic to the basic transition matrix

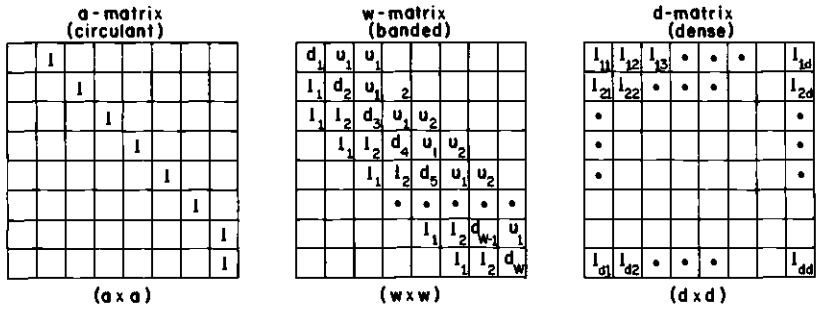


Figure 12.3 State transition matrices

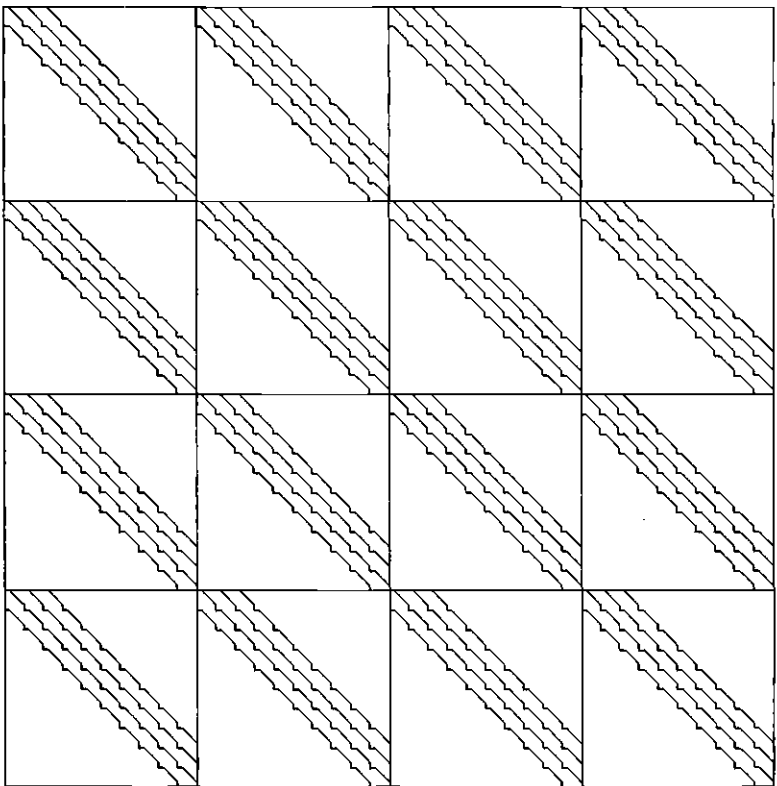


Figure 12.4 Sparsity pattern: (d, w, a)

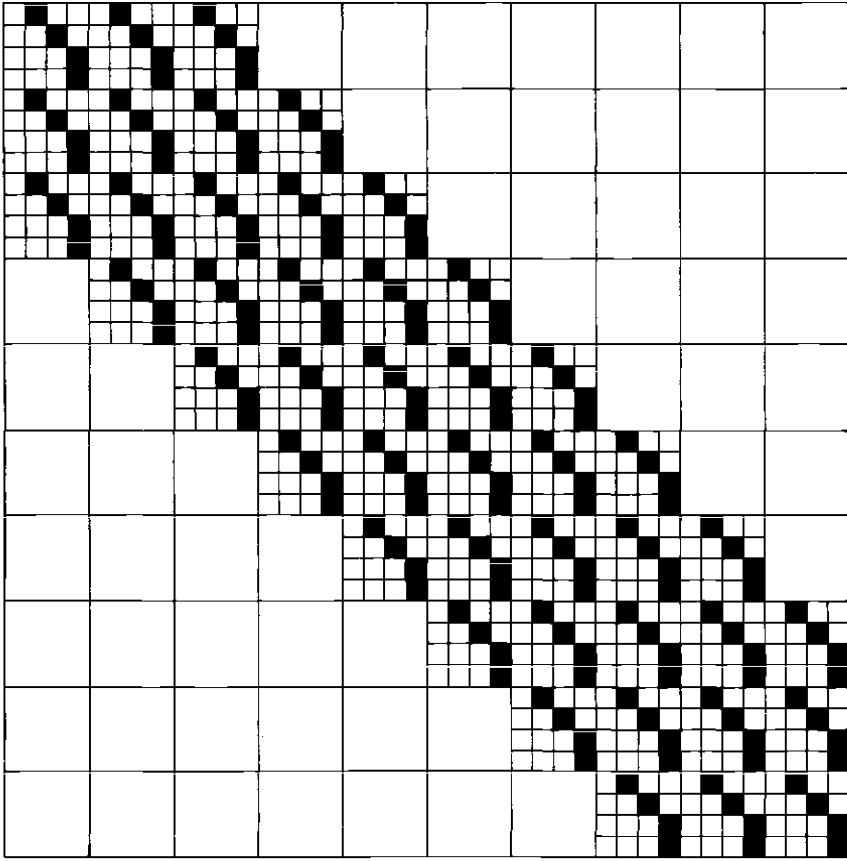


Figure 12.5 Sparsity pattern: (w,a,d)

for π . Thus, for each ordering of the underlying state variables, the matrix $[I - T'(v)]$ will have the same sparsity pattern as the matrix representation of π in figures 12.4 through 12.7 except for the 1's along the diagonal. Under the "optimal" ordering (a,w,d) , one can see from figure 12.8 that except for the lower (a,a) block, this matrix $[I - T'(v)]$ is already in upper triangular form. Thus, solving the linear system only requires an LU factorization of the lower (a,a) block followed by recursive back-substitution to compute the solution for age groups $a - 1$ to 1. Since LU factorization is an order N^3 operation, the time saved under the optimal ordering is proportional to a^3 , which amounts to a speedup of 8,000 times when $a = 20$. Further speedups can be obtained by accounting for the block-banded structure of the (a,a)

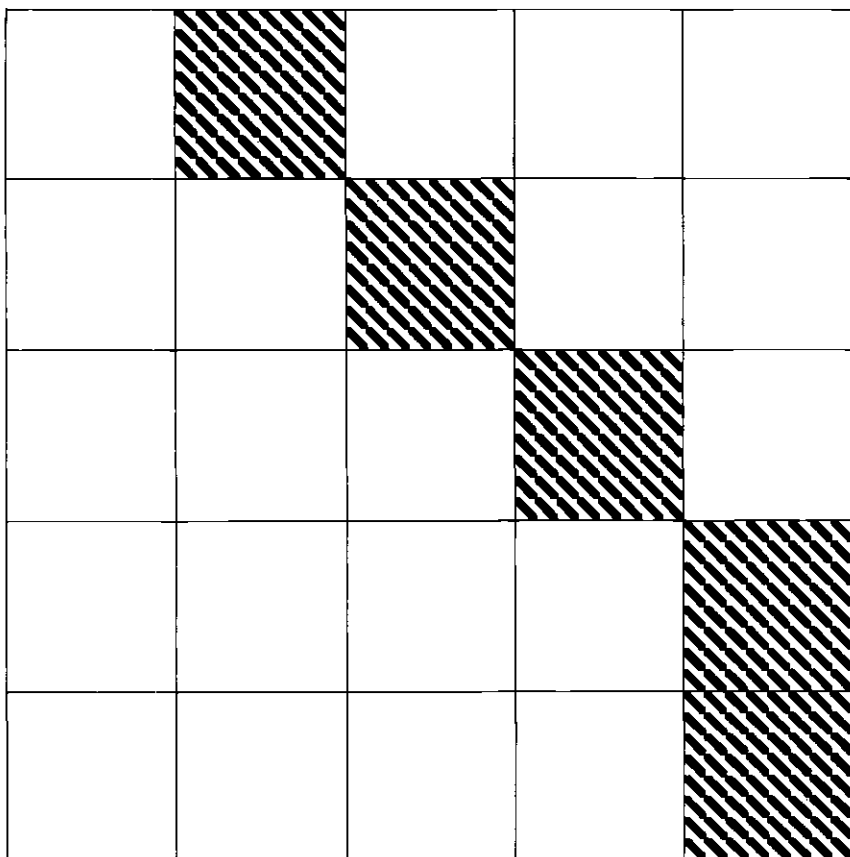


Figure 12.6 Sparsity pattern: (a, d, w)

block of $[I - T'(v)]$. I have designed a block elimination algorithm which LU factors the (a, a) block of $[I - T'(v)]$ using a banded-Crout decomposition, with elimination operations that are performed on $d \times d$ blocks instead of individual matrix elements. The matrix $[I - T'(v)]$ has sufficient diagonal dominance that the block elimination algorithm is numerically stable even though pivot operations only occur within the elementary $d \times d$ block operations of the block elimination procedure.⁹ Thus, by determining the optimal ordering of state variables one can design a special linear equation algorithm that fully exploits the sparsity structure of the $[I - T'(v)]$ matrix while keeping the vector processors running continuously at nearly peak efficiency. This fortuitous situation allows one to solve linear systems that are orders of

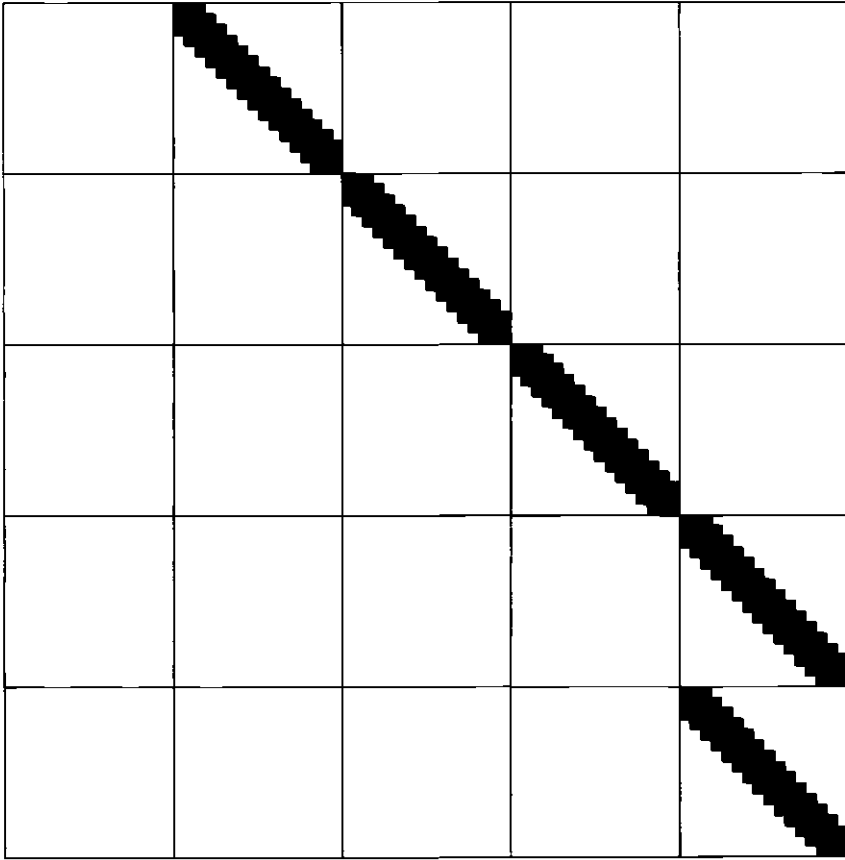


Figure 12.7 Sparsity pattern: (a, w, d)

magnitude larger than the largest systems solvable using standard sparse matrix software.

I conclude with table 12.6 which presents timings of the fixed point algorithm on the Cray-2. As one can see, the “finite horizon” assumption that workers die with probability 1 after age 98 allows one to expand the dimension of the problem by an order of magnitude. The average performance rate of 220 megaflops is good performance for a single processor bank of the Cray-2.¹⁰ Overall, table 12.6 demonstrates that one can exploit the power of the supercomputer and the special structure of the fixed-point problem to permit estimation of a fairly realistic model of retirement behavior. In future work I plan to use this technology to actually estimate the unknown parameters of the model.¹¹

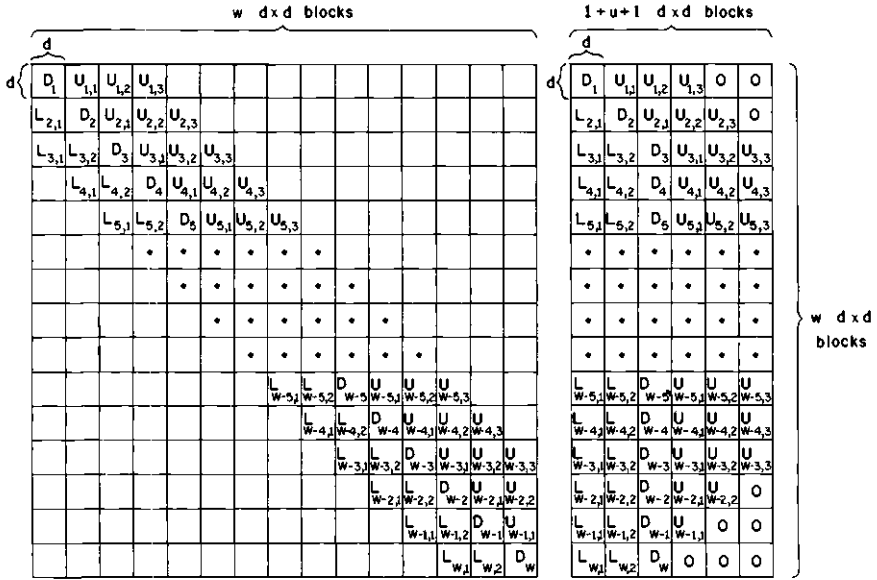


Figure 12.8 Packed storage form for block-banded matrix

Table 12.6 Fixed Point Computation Times on the Cray-2

Item	Infinite Horizon	Finite Horizon
Age categories, a	20	20
Wealth categories, w	100	100
Dense block size, d	50	90
Consumption levels, $\#c$	0	5
Labor decisions, $\#d$	3	3
Maximum bandwidth (blocks)	10	10
Fixed point dimension, N	300,000	2,700,000
CPU time (seconds)	14.3	6.9
Average rate (megaflops)	198	220

Note: Times are for one processor on the University of Minnesota four-processor Cray-2 with 256 million word common memory.

Notes

1. Burtless and Moffitt (1984) and Gustman and Steinmeier (1986) provided two of the most recent structural analyses of retirement behavior. Burtless and Moffitt allowed state-dependent preferences, but simplified the sequential decision problem by assuming that pre- and post-retirement consumption levels are fixed, leading to a two-stage approximation to the sequential labor supply-consumption decision. Gustman and Steinmeier estimated a fully sequential continuous-time model of consumption and labor supply, however, they assumed perfect certainty and perfect capital markets.

2. The authors reported, however, that "the correspondence among outcomes based on these alternative definitions (i.e. objective measures based on reductions in hours or wages) is relatively close, and the main conclusions of the paper remain unchanged using the alternative measure" (fn. 7, p. 405).

3. A statistical model is *degenerate* if a subset of its variables satisfy a relation with probability 1.

4. There have been recent advances in estimation methods for static discrete/continuous choice models by Dubin and McFadden (1984) and Haneman (1984). However, it is unclear whether these methods will extend to dynamic programming models. An alternative possibility is to attempt to merge the "orthogonality condition" method of Hansen and Singleton (1982) with the discrete choice framework of Rust (1986). A difficulty with this approach has been to specify a tractable stochastic process for the continuum of unobservables corresponding to each possible value of the continuous choice variable. Without such unobservables, one obtains a statistically degenerate model where the continuous choice variable is an exact function of other observed variables in the model.

5. The RHS has incomplete data on consumption expenditures. Rather than use this data directly, one can compute consumption from the budget equation $w_{t+1} = w_t + y_t - c_t$, since both income and wealth are measured much more completely and accurately in the RHS. An unfortunate complication is that the RHS records total income only for even-numbered years. Therefore, one must impute income in odd-numbered years based on retrospective information on labor force status in those years and a matched data file on Social Security earnings available for both even- and odd-numbered years.

6. These functional forms presented should be viewed as first guesses as to which specifications will "work." The final specification will be chosen from the results of a specification search over alternative functional forms using the NFXP algorithm.

7. Recent advances in simulation estimators by McFadden (1988) and Pakes and Pollard (1986) offer the hope of significantly extending the range of estimable distributions q for the unobservables. However, it is not clear whether their methods, which depend heavily on having the simulation errors enter linearly and additively separately, directly extends to allow simulation instead of integration in the fixed point condition (8). In that case the simulation error is no longer additively separable, so the simulations must increase with the sample size to avoid inconsistency due to nonlinear "errors in variables".

8. Rust (1989) provides estimation results for workers' expectations using specifications outlined in this section.

9. A simple check of the numerical accuracy of the method is to compute the fixed point with b and u identically equal to 1. It is easy to see that in this case $v_\theta = 1/(1 - \beta)$, so the numerical results of the algorithm can be checked against this exact solution. Running the algorithm in 64-bit single precision with $\beta = .999999$, I found that the computed solution agreed with the theoretical solution to twelve significant digits.

10. A simulation analysis of the Cray-2 processor (on a MacIntosh) by Lawrence Liddiard (1986) suggests that a single processor can achieve a maximum rate of 433 megaflops for dense matrix multiplication. The highest rates that have been recorded for the University of Minnesota Cray-2 (as of March 1987) have been on the order of 360 megaflops. The single processor average of 220 megaflops reported in table 12.6 has been achieved using standard library kernels without special assembly language coding to optimize the flow of data from common memory to local memory and to the vector registers, and back

to common memory. The easiest way to get significant speed increases is to utilize all four processors of the Cray-2 simultaneously. The fixed point computation can be fairly easily decomposed into sets of four independent subtasks (e.g., a separate processor is dedicated to computing the fixed point v_θ and the remaining processors assigned to computing each of the derivatives $\partial v_\theta / \partial \theta$), allowing a sustainable processing rate approaching 880 megaflops.

11. For estimates of workers' expectations, $\hat{\pi}$, see Rust (1989).

References

- Aaron, Henry J., and Gary Burtless. 1984. *Retirement and economic behavior*. Washington, D.C.: Brookings Institution.
- Anderson, K. H., R. V. Burkhauser, and J. S. Butler. 1988. Work and health after retirement: A semi-parametric bivariate interval hazard model. Manuscript, Vanderbilt University.
- Anderson, K. H., R. V. Burkhauser, and J. F. Quinn. 1984. Do retirement dreams come true? The effect of unexpected events on retirement age. Working Paper no. 750-84, Institute for Research on Poverty, University of Wisconsin-Madison.
- Barro, R. J. 1974. Are government bonds net worth? *Journal of Political Economy* 82(6):1095-1117.
- . 1978. *The impact of Social Security on private saving*. Washington, D.C.: The American Enterprise Institute.
- Bernheim, D. 1984. Dissaving after retirement: Testing the pure life cycle hypothesis. NBER Working Paper no. 1409. Cambridge, Mass.: National Bureau of Economic Research.
- Blackwell, D. 1965. Discounted dynamic programming. *Annals of Statistics* 36:226-35.
- Blinder, A., R. H. Gordon, and D. E. Wise. 1983. Social Security, bequests, and the life-cycle theory of savings. In *The determinants of national saving and wealth*, ed. F. Modigliani and R. Hemming, 89-122. New York: St. Martin's Press.
- Bodie, Z., and J. B. Shoven, eds. 1983. *Financial aspects of the United States pension system*. Chicago: University of Chicago Press.
- Boskin, M., and L. J. Kotlikoff. 1985. Public debt and United States savings: A new test of the neutrality hypothesis. In *The "new monetary economics," fiscal issues and unemployment*, ed. K. Brunner and A. H. Meltzer, vol. 23. Amsterdam: North-Holland.
- Boskin, M., L. J. Kotlikoff, D. J. Puffert, and J. B. Shoven. 1986. Social Security: A financial appraisal across and within generations. NBER Working Paper no. 1891. Cambridge, Mass.: National Bureau of Economic Research.
- Boskin, M., and M. Hurd. 1974. The effect of Social Security on early retirement. *Journal of Public Economics* 10:361-77.
- Boskin, M., and J. B. Shoven. 1986. Poverty among the elderly: Where are the holes in the safety net? NBER Working Paper no. 1923. Cambridge, Mass.: National Bureau of Economic Research.
- Bound, J. 1986. The disincentive effects of the Social Security disability program. University of Michigan. Typescript.
- Burkhauser, R. V. 1980. The early acceptance of Social Security: An asset maximization approach. *Industrial and Labor Relations Review* 33:484-92.

- _____. 1979. The pension acceptance decision of older workers. *Journal of Human Resources* 14(1):63–75.
- Burkhauser, R. V., and J. A. Turner. 1978. A time-series analysis of Social Security and its effect on the market work of men at younger ages. *Journal of Political Economy* 86:701–15.
- Burtless, Gary, and R. Moffitt. 1984. The effect of Social Security benefits on the labor supply of the aged. In *Retirement and economic behavior*, ed. Henry J. Aaron and Gary Burtless. Washington, D.C.: Brookings Institution.
- Cartwright, W. S., and R. B. Friedland. 1985. The President's Commission on Pension Policy Household Survey 1979. *Review of Income and Wealth* 31:285–308.
- Chirikos, T. N., and G. Nestel. 1981. Impairment and labor market outcomes: A cross-sectional and longitudinal analysis. In *Work and retirement*, ed. H. Parnes. Cambridge, Mass.: MIT Press.
- Congressional Quarterly, Inc. 1983. *Social Security and retirement*. Washington, D.C.: GPO.
- Crawford, V. P., and D. M. Lilien. 1982. Social Security and the retirement decision. *Quarterly Journal of Economics* (August):505–29.
- Danziger, S., R. Haveman, and R. Plotnick. 1981. How income transfers affect work, saving, and the income distribution. *Journal of Economic Literature* 19(3):975–1028.
- Danziger, S., J. van der Gaag, E. Smolensky, and M. Taussig. 1982. The life-cycle hypothesis and the consumption behavior of the elderly. *Journal of Post-Keynesian Economics* 5:208–27.
- Darby, M. R. 1979. *The effects of Social Security on income and the capital stock*. Washington, D.C.: American Enterprise Institute.
- Davies, J. B. 1981. Uncertain lifetime, consumption, and dissaving in retirement. *Journal of Political Economy* 89(3):561–77.
- Deaton, A. 1986. Life-cycle models of consumption: Is the evidence consistent with the theory? NBER Working Paper no. 1910. Cambridge, Mass.: National Bureau of Economic Research.
- Diamond, P. 1977. A framework for Social Security analysis. *Journal of Public Economics* 8(3):275–98.
- Diamond, P., and J. Hausman. 1984a. Individual retirement and savings behavior. *Journal of Public Economics* 23:81–114.
- _____. 1984b. The retirement and unemployment behavior of older men. In *Retirement and economic behavior*, ed. H. Aaron and G. Burtless. Washington, D.C.: Brookings Institution.
- Diamond, P., and J. A. Mirrlees. 1978. A model of social insurance with variable retirement. *Journal of Public Economics* 10:295–336.
- Dubin, J., and D. McFadden. 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* 52(2):345–62.
- Duff, I. S. 1984. The solution of sparse linear equations on the Cray-1. In *High speed computation*, ed. J. S. Kowalik. Berlin: Springer-Verlag.
- Feldstein, M. 1974. Social Security, induced retirement and aggregate capital accumulation. *Journal of Political Economy* 82:905–26.
- _____. 1983. Social Security benefits and the accumulation of pre-retirement wealth. In *The determinants of national wealth and saving*, ed. F. Modigliani and R. Hemming, 3–23. New York: St. Martin's Press.
- Feldstein, M., and A. J. Pellachio. 1979. Social Security and household wealth accumulation: New microeconomic evidence. *Review of Economics and Statistics* 61(3):361–68.
- Fields, G. S., and O. S. Mitchell. 1985. *Retirement, pensions, and Social Security*. Cambridge, Mass.: MIT Press.

- Fuchs, V. R. 1982. Time preference and health: An exploratory study. In *Economic aspects of health*, ed. V. Fuchs. Chicago: Univ. of Chicago Press.
- Gordon, R. H., and A. S. Blinder. 1980. Market wages, reservation wages, and retirement decisions. *Journal of Public Economics* 14:277–308.
- Gotz, G. A., and J. J. McCall. 1984. A dynamic retention model for Air Force officers. Research report R-3028-AF. Santa Monica, Calif.: The Rand Corporation.
- 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.
- Hall, A., and T. Johnson. 1980. The determinants of planned retirement age. *Industrial and Labor Relations Review* 33(2):241–55.
- Hamermesh, D. 1984. Consumption during retirement: The missing link in the life cycle. *Review of Economics and Statistics* 66:1–7.
- Haneman, W. M. 1984. Discrete/continuous models of consumer demand. *Econometrica* 52(3):541–62.
- Hansen, L. P., and K. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50(5):1269–86.
- Harris, L. 1984. *Retirement and income*. New York: Garland Publishing.
- Hurd, M. D. 1986. Savings and bequests. NBER Working Paper no. 1826. Cambridge, Mass.: National Bureau of Economic Research.
- Irelan, L., D. Motley, K. Schwab, S. Sherman, and J. Murray. 1976. *Almost 65: Baseline data for the retirement history study*. HEW Research Report 49. Washington, D.C.
- King, M. 1985. The economics of saving: A survey of recent contributions. In *Frontiers of economics*, ed. K. Arrow and S. Harkapohja. Oxford: Basil Blackwell.
- King, M., and L. D. Dicks-Mireaux. 1982. Asset holdings and the life-cycle. *The Economic Journal* 92:247–67.
- Kotlikoff, L. J. 1979. Testing the theory of Social Security and life cycle accumulation. *American Economic Review* 69:396–410.
- . 1986. Health expenditures and precautionary savings. NBER Working Paper no. 2008. Cambridge, Mass.: National Bureau of Economic Research.
- Kotlikoff, L. J., and D. E. Smith. 1983. *Pensions in the American economy*. Chicago: University of Chicago Press.
- Kotlikoff, L. J., and L. H. Summers. 1981. The role of intergenerational transfers in aggregate capital accumulation. *Journal of Political Economy* 89:706–32.
- Kotlikoff, L. J., A. Spivak, and L. H. Summers. 1982. The adequacy of savings. *American Economic Review* 72(5):1056–69.
- Kurz, M. 1981. The life-cycle hypothesis and the effects of Social Security and private pensions on family savings. IMSSS Technical Report 335, Stanford University.
- . 1984. Capital accumulation and the characteristics of private intergenerational transfers. *Economica* 51: 1–22.
- . 1985. Heterogeneity and savings behavior: A comment. In *Frontiers of economics*, ed. K. Arrow and S. Harkapohja. Oxford: Basil Blackwell.
- Levhari, D., and L. J. Mirman. 1977. Savings and consumption with an uncertain horizon. *Journal of Political Economy* 85(2):265–81.

- Liddiard, L. D. 1986. How to get the most out of a Cray-2. University of Minnesota. Typescript.
- Lucas, R. E. 1976. Econometric policy evaluation: A critique. In *The Philips Curve and labor markets*, ed. K. Brunner and A. K. Meltzer. Carnegie-Rochester Conference on Public Policy 1. Amsterdam: North-Holland.
- 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.
- Marschak, J. 1953. Economic measurements for policy and prediction. In *Studies in econometric method*, ed. W. C. Hood and T. C. Koopmans. New York: Wiley and Sons.
- McFadden, D. 1981. Econometric models of probabilistic choice. In *Structural analysis of discrete data with econometric applications*, ed. C. F. Manski and D. McFadden. Cambridge: MIT Press.
- _____. 1988. A method of simulated moments for estimation of multinomial probits without numerical integration. *Econometrica*. Forthcoming.
- Menchik, P. L., and M. David. 1983. Income distribution, lifetime savings, and bequests. *American Economic Review* 73(4):672-90.
- _____. 1985. The effect of Social Security on lifetime wealth accumulation and bequests. *Economica* 52(208):421-34.
- Miller, R. 1984. Job matching and occupational choice. *Journal of Political Economy* 92(6):1086-1120.
- Mirer, T. W. 1979. The wealth-age relation among the aged. *American Economic Review* 69(3):435-43.
- _____. 1980. The dissaving behavior of the aged. *Southern Economic Journal* 46:1197-1205.
- Mitchell, O. S., and G. S. Fields. 1984. The economics of retirement behavior. *Journal of Labor Economics* 2(1):84-105.
- Modigliani, F. 1975. The life cycle hypothesis of saving twenty years later. In *Contemporary issues in economics*, ed. M. Pakin. Manchester: Manchester University Press.
- Modigliani, F., and F. Brumberg. 1954. Utility analysis and the consumption function: An interpretation of cross-section data. In *Post-Keynesian economics*, ed. K. Kurihara. New Brunswick: Rutgers University Press.
- Morrison, M. H. 1982. *Economics of aging: The future of retirement*. New York: Van Nostrand Reinhold.
- Mott, F. L., and R. J. Haurin. 1981. The impact of health problems and mortality on family well-being. In *Work and retirement*, ed. H. Parnes. Cambridge: MIT Press.
- Munnell, A. H. 1977. *The future of Social Security*. Washington, D.C.: Brookings Institution.
- Pakes, A. 1986. Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica* 54(4):755-84.
- Pakes, A., and D. Pollard. 1988. The asymptotics of simulation estimators. *Econometrica*, forthcoming.
- Parnes, H. S., ed. 1981. *Work and retirement*. Cambridge: MIT Press.
- _____, ed. 1983. *Policy issues in work and retirement*. Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.
- Parsons, D. O. 1982. The male labour force participation decision: Health, reported health, and economic incentives. *Economica* 49: 81-91.
- Pellechio, A. J. 1979. Social Security financing and retirement behavior. *American Economic Review* 69(2):284-87.

- Quinn, J. F. 1977. Microeconomic determinants of early retirement: A cross-sectional view of white married men. *Journal of Human Resources* 12: 329–46.
- Quinn, J. F., and R. V. Burkhauser. 1983. Is mandatory retirement overrated?: Evidence from the 1970's. *Journal of Human Resources* 18(3):377–58.
- Rust, J. 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55(5):999–1033.
- . 1988. Maximum likelihood estimation of discrete control processes. *SIAM Journal on Control and Optimization* 26(5): 1–19.
- . 1989. Behavior of male workers at the end of the life-cycle: An empirical analysis of states and controls. In *Issues in the economics of aging*, ed. D. Wise. Forthcoming.
- Sargent, T. J. 1981. Interpreting economic time series. *Journal of Political Economy* 89(2):213–48.
- Schulz, J. H. 1976. *The economics of aging*. Belmont, Calif.: Wadsworth Publishing.
- Sickles, R. C., and P. Taubman. 1986. An analysis of the health and retirement status of the elderly. *Econometrica* 54(6):1339–56.
- Sheshinski, E. 1978. A model of Social Security and retirement decisions. *Journal of Public Economics* 10: 337–60.
- Stein, B. 1980. *Social Security and pensions in transition*. New York: Free Press.
- Sueyoshi, G. 1986. Social Security and the determinants of full and partial retirement: A competing risks analysis. M.I.T. Typescript.
- Thompson, L. H. 1981. The Social Security reform debate. *Journal of Economic Literature* 21(4):1425–67.
- Tobin, J. 1967. Life-cycle saving and balanced growth. In *Ten economic studies in the tradition of Irving Fisher*, ed. W. Fellner, 231–56. New York: Wiley and Sons.
- White, B. 1978. Empirical tests of the life-cycle hypothesis. *American Economic Review* 68: 547–60.
- . 1984. Empirical tests of the life-cycle hypothesis: Reply. *American Economic Review* 74: 258–59.
- Williams, C. A., Jr., J. G. Turnbull, and E. F. Cheit. 1982. *Economic and social security: Social insurance and other approaches*. New York: Wiley and Sons.
- Wise, D., ed. 1985. *Pensions, labor, and individual choice*. Chicago: University of Chicago Press.
- Wolpin, K. 1984. An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy* 92(5):852–74.
- Yaari, M. E. 1965. Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies* 32: 137–50.
- Zabalza, A., C. Pissarides, and M. Barton. 1980. Social Security and the choice between full-time work, part-time work, and retirement. *Journal of Public Economics* 14: 245–76.

Comment Gary Burtless

The author of this paper sets out to do a couple of notable things. First, he wants to set up a tractable “optimizing” model that workers follow

Gary Burtless is a senior fellow of the Brookings Institution.

in determining their period-by-period consumption and labor supply toward the end of life. Second, he attempts to embed this model within a defensible econometric framework so that the important unknown parameters in the model can be estimated with a longitudinal data base containing information on several thousand men. Third, he hopes to obtain empirical estimates of the unknown parameters within our lifetime and within a very, very large—although finite—computer budget. And finally, I guess, he would eventually like to describe his model and results in a paper that is accessible to an audience that includes, but is not *restricted* to, operations research specialists who are conversant with the internal architecture of the Cray-2 supercomputer.

Judging by the evidence in this paper, the last objective might turn out to be the hardest to attain.

You might ask: How does this set of objectives distinguish this paper from other papers on the subject? To my knowledge, no previous analyst has ever been quite so ambitious. Some analysts, including myself, have examined a *subset* of the issues treated in this paper: What determines the age at retirement? What determines an older worker's weekly hours? How can we explain the dynamic pattern of partial and full retirement? What are the determinants of the saving/consumption path, particularly the trends in saving before and after retirement? How do unexpected events, such as layoffs, poor health, and death, affect labor supply and consumption?

Many previous analysts—also including myself—make a ritual bow toward the idea that each of these issues can and should be treated within a unified life-cycle, labor supply/consumption model. But then in the empirical implementation we plead lack of data or an insufficient computer budget. We treat only one or occasionally two of the issues in the list I just mentioned. This author is after the whole nine yards: A unified framework that simultaneously treats *all* of the issues.

Devising such a framework is very, very hard. Before talking about the author's proposed solution, I would like to describe just how hard it is. In the canonical model of labor supply and consumption, economic agents are assumed to make decisions about current and future hours of work and saving based on current and future wage rates, current and anticipated wealth—including wealth accumulated through Social Security and pensions—current and anticipated health and family responsibilities, and individual tastes—specifically, the taste for work versus leisure, leisure versus goods consumption, and present versus future consumption and leisure.

To add to the analyst's problem, he or she is confronted with a very limited, and in some cases defective, data set. Frequently, the current wage rate is poorly measured. Even where it is well measured, we have no way of knowing the worker's potential future wages. Wealth is notoriously badly measured. Even where it is well measured, it is

very tricky to provide simple conceptual definitions of future wealth, because such critical items as pension assets and Social Security accumulations are highly dependent on the exact sequence of potential wage rates and actual labor force participation decisions. (For example, wealth accumulation in Social Security occurs in a highly nonlinear way as the retirement age is delayed from 59 to 62, from 62 to 65, and from 65 to 68.)

Even if we can solve these measurement and conceptual problems in obtaining observations on a worker's wealth, there is still the problem that the wealth concept measured in the data usually does not represent the worker's initial endowment, which is, of course, the correct starting point for measuring his or her lifetime budget constraint. Instead, we observe wealth at some initial age (in this case, ages 58 to 63, which are the earliest ages observed in the Retirement History Survey). Practically speaking, wealth at those ages probably reflects most of the assets the worker is going to have available to finance his retirement. Moreover, those observed wealth levels embody all the worker's past consumption and labor supply decisions, given his initial endowment and potential market wage. In terms of the full life-cycle model, the observed wealth holdings at the start of the survey represent endogenous—not exogenous—variables.

If these measurement problems are not enough to deter even a stout-hearted graduate student seeking econometric glory, let me add one final point. In the canonical life-cycle model we are not really talking about a straightforward two-good problem—goods consumption and leisure. At a minimum we are talking about four goods—more probably $2 \times T$, where T is the expected number of periods in the lifetime. That is, we are talking about consumption in the current period and in a future period, and leisure in the current period and in a future period. The large number of goods in the utility function gives rise to an integrability problem in deriving individual demand functions, except in forms of the function that are carefully specified a priori, and which might be unrealistic.

Armed with a guarantee of ten hours on a Cray-2, John Rust bravely confronts these problems and outlines a research program. His courage and high intelligence are everywhere evident in this paper. But, I would need a better weapon than mere CPU time before hazarding into this particular battle.

How does Rust overcome the problems I just mentioned? It is not clear that some of them are even treated. For example, I am not sure that Rust considers that the potential wage of workers might vary over time. He proposes to enter the wage or potential earnings of a worker at each conceivable level of work effort (here taken to be zero hours, part-time hours, and full-time hours) as explanatory variables. But does

he account for the fact that the potential market wage might decline with advancing age? Or for the fact that the tax on wages must change at different ages, depending on acceptance or rejection of a Social Security pension? I do not know. But this is a central problem traditionally faced in the retirement literature, and it does not appear to be mentioned here.

How does the author treat the issue of nonlinear accumulation of pension and Social Security wealth? (These accumulations are nonlinear with respect to the age of retirement because of minimum pension ages, early retirement penalties, and actuarially unfair compensation for workers delaying retirement beyond the normal retirement age.) The author handles the problem posed by private pensions by tossing out every member of his sample who is covered by one. He handles Social Security accumulations by assuming that the remaining members of his sample perfectly anticipated in 1969 and 1971 that Social Security wealth would be raised by the 1972 amendments. That is, he assumes that all workers plan on receiving benefits as Social Security pensions would be computed under the 1973 law. These people are indeed farsighted.

Let me interject here that this measure of Social Security wealth is extremely troubling. In this paper we have a general model that must account for all anticipated and unanticipated events that might affect retirement and consumption. Yet the author fails to use our best available information about when the workers learned of Social Security benefit increases; namely, the date that real benefits were raised by 8–10 percent in late 1969 and a further 9–11 percent in late 1972.

Although I intuitively understand how the author will handle the growth in Social Security wealth that can be anticipated, I am not really sure how he will do it, because that issue is not explicitly treated in the paper.

How does the author handle the endogeneity of his initial observation on wealth? Here he cuts the Gordian knot and simply claims he does not *need* to address the problem. All he needs to do is estimate parameters that generate transition probabilities between the many specified statuses, given an initial set of state variables, including initial wealth. I find this claim difficult to reconcile with another objective of the paper. Rust wants to use the estimated model to forecast the effect of policy changes, such as reductions in Social Security retirement or disability benefits. This implies that the parameter estimates recovered from the analysis can be used to forecast the effect of policy changes on any future cohort of men near retirement age.

Suppose we think of a sample of men whose wealth accumulations occurred under a radically different sequence of state variables: different Social Security and income tax rates or different real interest

rates and patterns of capital gain. Then as I understand the paper, we are being told that the transition probabilities (i.e., the underlying parameters) estimated in this study can be applied to this new sample. Personally, I find this hard to believe, but I am willing to be persuaded.

I will not pretend that I understand how the author expects to estimate this model. It would be too easy for the alert reader to detect my ignorance of dynamic programming techniques. But I will open myself up to ridicule by admitting that I do not quite understand how he intends to organize his data and define an observation, e.g., a dependent variable. The paper lays great stress on a worker's decision about making an optimal plan at a given point in time. Does this mean that the author will rely on the worker's stated intentions to measure projected labor supply and consumption paths? That is, in defining the "state" variables presumed to exist at a given point in time, he will rely on the worker's observed employment status and some direct estimate of his recent consumption level. But how will the author measure the worker's "choice" in that period? By using the observed labor market and consumption path over the next two years? Or by using the worker's stated intention at that time to retire at a specified date in the future? Both of these options seem to me to have advantages and problems, but it is not entirely clear which option has been selected.

The reader should not interpret my comments as being especially critical. Many of the questions I raise will undoubtedly be cleared up in a subsequent paper, in which the estimation results are actually displayed. Other points I appear to criticize are probably necessary simplifications in order to make estimation possible. Anyone working in this area has had to make the same kinds of more or less defensible simplifications.

Some readers might feel that using the word "simplification" anywhere within ten miles of this paper is a little like using the word "integrity" in discussing the ethical standards of the advertising industry. But it is nonetheless the case that Rust has had to make a certain number of drastically simplifying assumptions to make estimation feasible. He will probably have to make even more before obtaining estimates.

Even with these simplifications, however, we might wonder whether the final result is simple enough to reflect the prosaic reality that surrounds us. In thinking about consumer decision models, I like to apply the "Paul Pepperbridge" test. Paul Pepperbridge was the average graduate of the average high school in the United States. If a decision model seems to call for more intellectual capacity than Paul can bring to bear, I am usually a little skeptical. If it calls for a couple of seconds of CPU time on a Cray-2 to assess all the consumption alternatives, I am a bit more skeptical. Let's face it, the Cray can probably make as many

calculations in a second as Paul was willing and capable of making between the sixth and twelfth grades. I doubt he will spend much more time than that thinking about when he is going to retire.

So one thing the author might want to do is persuade the reader that the decision-making model provides a conceivable approximation to reality, as well as the best a consumer could do if he had a couple of hours on a Cray and the professional advice of a consulting economist. From the point of view of a fellow economist, however, the model described in this paper represents an extremely impressive achievement.

