NBER WORKING PAPER SERIES

# A DYNAMIC PROGRAMMING MODEL OF RETIREMENT BEHAVIOR

John Rust

Working Paper No. 2470

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1987

Forthcoming in David Wise, (ed), The Economics of Aging, NBER, Cambridge, MA 1987. The research reported here is part of the NBER's research project on Aging. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #2470
December 1987

A Dynamic Programming Model of Retirement Behavior

ABSTRACT

This paper formulates a model of retirement behavior based on 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 the worker chooses how much to consume and whether to work full-time, part-time, or exit the labor force. The model accounts for the sequential nature of the retirement decision problem, and the role of expectations of uncertain future variables such as the worker's future lifespan, health status, marital and family status, employment status, as well as earnings from employment, assets, and social security retirement, disability and medicare payments. This paper applies a "nested fixed point" algorithm that converts the dynamic programming problem into the problem of repeatedly recomputing the fixed point to a contraction mapping operator as a subroutine of a standard nonlinear maximum likelihood program. The goal of the paper is to demonstrate that a fairly complex and realistic formulation of the retirement problem can be estimated using this algorithm and a current generation supercomputer, the Cray-2.

Professor John Rust
Department of Economics
University of Wisconsin
Madison, WI 53706

## 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 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 and 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 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 $\pi$ that depend on a vector of unknown parameters $\theta$. The underlying preferences u and expectations $\pi$ 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 appears 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 (1957) 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 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 (1986) 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 sucessfully 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 2 reviews some of the empirical issues that motivated the construction of the model. Section 3 develops the model, formulating the retirement decision process as a discrete control process. Section 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. A future paper will use the NFXP algorithm and data from the longitudinal Retirement History Survey (RHS) to actually estimate the unknown parameters of the model.

## 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 arisen over the last 20 years. This section summarizes some of the basic empirical and policy issues of the retirement process that I wanted the model to capture.

### 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 pre-retirement and post-retirement phase. At some initial "planning date" before retirement, the worker is assumed to choose a fixed optimal retirement date and fixed pre-retirement and post-retirement 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 non-retired workers reported their planned retirement age. By tracing workers over the subsequent 10 years they were able to compare the actual and planned retirement dates, and found that over 40% 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.

## 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 (Mirer (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) page 224). A study of consumption profiles using the RHS data by Hammermesh (1984) found that on average consumption exceeds earnings by 14% 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" (page 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 U.S., 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 6 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 1970's are predicted to be completely offset by increases in private savings 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 (1984), 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% 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 seems to be simply the result of mortality risk combined with a very weak market for private annuities" (page 35). David and Menchik's (1985) study also casts doubt on 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. David and Menchik 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." (page 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.

## 2.3 Accounting for the joint endogeneity of labor supply and savings decisions

The decline in labor force participation rate of older males over the past 30 years is a well-known phenomenon; the participation rate for workers aged 55-64 declined from 86.8% in 1960 to 72.3% in 1980, and the rate for workers aged 65+ declined from 33.1% to 19.1% 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% 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% in the 50's, 8.7% in the 60's, 7.7% in the 70's, to only 5.1% since 1980. Some researchers including Feldstein (1974) have claimed that social security "depresses personal saving by 30-50 percent" (Feldstein (1974), page 905). However according to economic theory an actuarially fair social security program should have no effect on aggregate savings or labor supply decisions, 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, with 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% 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." (page 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 won't 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.

## 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% reported they were retired due to poor health; for those who had been out of the labor force for more than 6 years (the early retirees) the figure was 82%. Health problems are prevalent even among those who work; 39% of the 1969 RHS sample reported a health problem that limited their ability to work or get around, even though 63% 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 lead 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 1) that suggests that much of the decline in the labor force participation rates of older males over the last 30 years can be ascribed to increases in disability claims allowed under the social security disability insurance program instituted in the late 50's and substantially liberalized during the 70's. 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.

(figure 1 here)

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), page 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 doesn't 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. social retirement benefits on the declining labor force participation rate of older males.

## 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 hour 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 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 1 presents the sequence of self-reported labor market states in the first 4 waves of the RHS.

(table 1 here)

Table 1 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. Table 1 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 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.

(table 2 here)

A possible explanation for the discrepancy is that Gustman and Steinmeier used a self-reported measure of labor force status to construct Table 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 prohibit 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 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), page 277). The next section presents a model and estimation algorithm that can accomodate 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.

## 3. Theoretical 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 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.

### 3.1 State and control variables

In order to represent 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 eligibilty 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% between 1968 and 1979, the majority of the increase, 46.7%, 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% tax on earnings beyond the previous earnings limit to a 50% 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 1969-1973, but maintained static expectations that no further changes would occur thereafter.

Given these expectations, at each time t the worker 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 worker's 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 his remaining lifetime, where his 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 towards 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 3.

(table 3 here)

### 3.2 Formulating retirement behavior as 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. 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 $\varepsilon_t$ that the econometrician doesn't observe. By incorporating such *unobserved state variables* one obtains a non-degenerate, internally consistent statistical model generated by optimal decision rules of the form $i_t=f_t(x_t,\varepsilon_t)$. Rust (1986) developed a formal statistical framework for structural estimation of discrete markovian decision problems with unobserved state variables. The following table summarizes the basic structure of the problem:

(table 4 here)

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

$$(3.1) \qquad V(x_t,\varepsilon_t) = \sup_{\Pi} E\{\sum_{j=t}^{\infty} \beta^{(j-t)}[u(x_j,f_j)+\varepsilon_j(f_j)]|x_t,\varepsilon_t\}$$

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 $\varepsilon_t$, and where the expectation is taken with respect to the transition density for the controlled stochastic process $\{x_t,\varepsilon_t\}$ determined from $\Pi$ and the transition density $p(x_{t+1},\varepsilon_{t+1}|x_t,\varepsilon_t,i)$. Under general conditions specified in Rust (1986), the value function V will be the unique solution to *Bellman's equation*

$$(3.2) \quad V(x,\varepsilon) = \max_{i\in C(x)} [u(x,i)+\varepsilon(i)+\beta EV(x,\varepsilon,i)]$$

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

$$(3.3) \quad EV(x,\varepsilon,i) \equiv \int_y \int_\eta V(y,\eta)p(dy,d\eta|x,\varepsilon,i).$$

Blackwell's Theorem implies that the solution $\Pi$ is *stationary* $\Pi=\{f,f,f,...\}$, 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

$$(3.4) \quad f(x,\varepsilon) = \underset{i\varepsilon C(x)}{\operatorname{argmax}} [u(x,i)+\varepsilon(i)+\beta EV(x,\varepsilon,i)].$$

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 $\varepsilon$ using the conditional density of $\varepsilon$ given x. From equation (3.4) 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, $\varepsilon_t$ will be continously 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 $\varepsilon$ in the optimal decision rule defined by (3.4). However, the expected value function $EV(x,\varepsilon,i)$ entering (3.4) will almost never have a convenient analytic formula, but must be computed by numerically integrating the value function V in (3.3). The value function must in turn be numerically computed by solving V as a functional fixed point to Bellman's equation (3.2). Since $\varepsilon$ 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 $\varepsilon_t$, the dimensionality of the resulting discrete approximation will generally be too large to be computationally tractable. These computational problems motivated Rust (1986) to make the following assumption on the joint transition density for $\{x_t,\varepsilon_t\}$:

*Conditional Independence Assumption:* The markov transition density factors as

$$p(x_{t+1},\varepsilon_{t+1}|x_t,\varepsilon_t,i)=q(\varepsilon_{t+1}|x_{t+1})\pi(x_{t+1}|x_t,i), \qquad i\in C(x_t). \tag{3.5}$$

This assumption involves two restrictions. First, $x_{t+1}$ is a sufficient statistic for $\varepsilon_{t+1}$, which implies that any statistical dependence between $\varepsilon_t$ and $\varepsilon_{t+1}$ 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 $\varepsilon_t$. Although (3.5) is a strong assumption, Rust (1987) developed a simple LaGrange multiplier statistic to test its validity. The payoff to assumption (3.5) is given by the following theorems of Rust (1986).

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

$$(3.6) \qquad G(v(x)|x) \equiv \int_{\varepsilon} \max_{i \in C(x)} [v(x,i)+\varepsilon(i)] q(d\varepsilon|x)$$

and let $G_i(v(x)|x)$ denote the partial derivative of $G(v(x)|x)$ with respect to $v(x,i)$. Then under assumption (3.5)[3] the conditional choice probability $P(i|x)$ is given by

$$(3.7) \qquad P(i|x)=G_i(v(x)|x) \qquad i \in C(x)$$

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

$$(3.8) \qquad v(x,i) = u(x,i)+\beta \int_y G(v(y)|y)\pi(dy|x,i) \quad i \in C(x).$$

The function $v$ is related to the value function $V$ defined in (3.1) and (3.2) by

$$(3.9) \qquad V(x,\varepsilon) = \max_{i \varepsilon C(x)} [v(x,i)+\varepsilon(i)].$$

Theorem 2: Under assumption (3.5) the controlled stochastic process $\{i_t,x_t\}$ is markovian with transition density given by

$$(3.10) \qquad Pr\{i_{t+1},x_{t+1}|i_t,x_t\}=P(i_{t+1}|x_{t+1})\pi(x_{t+1}|x_t,i_t)$$

Products of the transition density given in (3.10) 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 (3.8). Theorem 1 shows that from the standpoint of evaluating (3.10), there are two major payoffs to assumption (3.5). First, it implies that $\varepsilon$ does not enter the expected value function $EV(x,\varepsilon,i)$, so that $\varepsilon$ enters $V$ only additively as shown in (3.9). 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 (3.8). Second, assumption (3.5) 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 $\nabla=\{(x,\varepsilon)|x\in R^M, \varepsilon\in R^N, N=\#C(x)\}$ since $\varepsilon$ is a continuous-valued N-dimensional vector which must be discretized into $K^N$ values (where $K$ is the diameter of the grid for $\varepsilon$), 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 $\pi$, one can "recover" the agent's underlying preferences $(\beta,u)$ and expectations $(\pi,q)$ by finding parameter values that maximize the likelihood function. This suggests the following "nested fixed point algorithm": an "outer" non-linear optimization algorithm searches for parameter vector $\theta$ that maximizes the likelihood function, and an "inner" fixed point algorithm re-calculates the fixed point $v_\theta$ of (3.8) each time the outer optimization algorithm updates its estimate of $\theta$. Rust (1986) 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 $\pi$, 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 apology is that the discrete formulation seems to be the best available compromise 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 at time t, and the decisions $i_t=(d_t,c_t)$ refers 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 unforseen future events that cause actual consumption to deviate from the plan. The use of consumption intervals also helps 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 consumption intervals are sufficiently wide that the ex post realized consumption interval coincides with the ex ante plan.

## 3.3 Specification of Workers' Preferences

The following table summarizes the formulation of the retirement problem as a discrete control process.

(table 5 here)

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 time 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, $\pi$, and q.[6] The NFXP algorithm places no restrictions on the functional forms for b, u, and $\pi$ but computational tractability appears to require that the distribution of unobservables q be a member of McFadden's (1981) "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 (3.6) and its derivatives, the choice probabilities (3.7). 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

$$(3.11) \qquad Q(\varepsilon,\theta_4) = \exp\{-\sum_{\delta=1}^{3} [\sum_{j=1}^{J} \exp\{\varepsilon(\delta,j)/\theta_{4\delta}\}]^{\theta_{4\delta}}\}, \quad 0\le\theta_{4\delta}\le 0, \quad \delta=1,2,3.$$

Since the corresponding density q does not depend on x, it follows that the unobserved state variables are serially independent in this specification. Formula (3.11) 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 $\varepsilon_t$ represented by the following *choice tree*

(figure 2 here)

Thus, (3.11) 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 (3.11) yields the following *nested logit* formulas for the conditional choice probabilities

$$(3.12) \qquad P(d,c|x,\theta)=P(c|x,d,\theta)P(d|x,\theta)$$

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

$$(3.13) \qquad P(c|x,d,\theta) = \frac{\exp\{v_\theta(x,c,d)/\theta_{4d}\}}{\sum_{j=1}^{J} \exp\{v_\theta(x,j,d)/\theta_{4d}\}}$$

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

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

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

Finally (3.11) yields an explicit formula for the fixed point condition (3.8)

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

with the implicit "terminal condition" that $v_\theta(x,d,c)=b(w,ms,\theta_1)$ if hs="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

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

The coefficient $\theta_{11}$ will reflect a diminishing or increasing marginal utility of bequests depending on whether $\theta_{11}$ is greater or less than 1. The David and Menchik study discussed in section 2.3 suggests that possibly $\theta_{11}>1$. Presumably a married worker obtains greater utility from bequests to remaining spouse than from bequests to friends, institutions, or the government. Thus, I expect that $\theta_{13}$ is positive.

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

$$(3.18) \quad u(d,c,e,h,a,ms,\theta_2) = \left[\sum_{i=1}^{3}\sum_{j=1}^{3}\theta_{2,i*j} I\{d=i,e=j\}\right]\left[c^{\theta_{210}}\right]\left[e^{\theta_{211}}\right]\left[\theta_{212}+\theta_{213}a+\theta_{214}ms+\theta_{215}hs\right]$$

According to (3.18), utility is a function of consumption, c, and the level of leisure, e. Ranking the employment states as 1=ft, 2=pt, and 3=ne, I expect the coefficient $\theta_{211}$ should be negative for a leisure lover, and positive for a "workaholic". The coefficient $\theta_{210}$ 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 (3.18) 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 $\theta_{215}$ 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 $\theta_{213}$ 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 would increase the worker's utility, suggesting a positive value for $\theta_{214}$.

The final term in (3.18) 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 find 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 1 of section 2.5 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 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 do 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 socio-demographic variables, the most important of which are race and the worker's main career occupation and industry.

### 3.4 Specification of workers' expectations

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 income and his 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 $\pi$) within each socio-demographic stratum.

Since the transition probability $\pi$ only depends on observable variables $\{i_t, x_t\}$, one could in principle use non-parametric methods to estimate it. With a discrete state space, the non-parametric 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 non-parametric estimate of $\pi$ 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 $\pi$ 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 don't 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 $\pi$ is out of the question. It is much simpler to decompose $\pi$ 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 $x_t(m)$ 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 4. Therefore I assume that life ends with probability 1 at age 98 or before, implying that $a_t$ takes on 20 values in increments of 2 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 to limit his ability to get around or 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, 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 {1,2,3} obey a parametric transition probability of the form

$$\pi(h_{t+1}|h_t,a_t,ms_t,e_t,w_t,c_t,d_t;\theta_{31}) \tag{3.19}$$

which gives the probability of health next period as a function of health this period, age, marital status, employment status, wealth, and the labor supply and consumption decisions. The function $\pi$ 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. To see this, note that the conditional probability that $h_{t+1}=3$ given $h_t=2$ can be interpreted as a 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 $ms_t$, worker marital status, 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 {ft,pt,ne}, 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

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

It is particularly important to allow for the effects of age and health on re-employment probabilities. Wealth and income are included as proxys 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 that 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 workers' chance of being fired from his current job to increase with age and poor health, to decrease with "experience" as proxied by his current income and wealth, $y_t$ and $w_t$. I also expect that full-time jobs to be 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 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, income from investments, less consmuption expenditures:

(3.21) $w_{t+1}=w_t+y_t-c_t$

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

$$(3.22) \quad \pi(y_{t+1}|y_t, w_t, aw_t, h_t, a_t, e_t, ms_t, d_t, \theta_{33})$$

Here the function $\pi$ 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 ($e_t$=ft), then $\pi$ will predict his next period earnings on his job. These earnings will be a function of his age and 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. $\pi$ also includes investment income on existing wealth, $\tilde{r}w_t$, where $\tilde{r}$ 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 ($d_t$=pt or $d_t$=ft, and $e_t$=ne), then the earnings function predicts the worker's UI benefits. If the worker is retired, ($e_t$=ne and $a_t \geq 62$), then $\pi$ 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, $ms_t$. $\pi$ 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 $\pi$ 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 approriately altering the earnings function $\pi$ 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, and to 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.

## 4. Numerical computation of the dynamic programming model

As described in section 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, $\theta_1$ are the parameters entering the bequest function b, $\theta_2$ are the parameters entering the utility function u, $\theta_3$ are the parameters of the transition probability for the observed state variables $\pi$, and $\theta_4$ are the parameters of the transition probability for the unobserved state variables q. The unknown parameters can be estimated be maximum likelihood method using the following three step procedure:

Step 1 Estimate the vector $\theta_3$ 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

$$(4.1) \qquad L_1(\theta_3) \equiv \prod_{k=1}^{K} \prod_{t=1}^{5} \pi(x_{t+1,k}|x_{t,k},i_{t,k},\theta_3)$$

(where k indexes individuals in the RHS sample)

Step 2 Using the initial consistent estimate $\hat{\theta}_3$ 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

$$(4.2) \qquad L_2(\beta,\theta_1,\theta_2,\hat{\theta}_3,\theta_4) \equiv \prod_{k=1}^{K} \prod_{t=1}^{5} P(i_{t,k}|x_{t,k},(\beta,\theta_1,\theta_2,\hat{\theta}_3,\theta_4))$$

where P is defined by (3.12)-(3.15) and the fixed point condition (3.16)

Step 3 To get correct estimated standard errors and asymptotically efficient parameter estimates for $\theta$, compute 1 Newton-step from the initial consistent estimate $\hat{\theta}$ using the full likelihood function $L_f(\theta)$ defined by

$$(4.3) \qquad L_f(\theta) \equiv \prod_{k=1}^{K} \prod_{t=1}^{5} P(i_{t+1,k}|x_{t+1,k},\theta)\pi(x_{t+1,k}|x_{t,k},i_{t,k},\theta_3)$$

The nested fixed point algorithm is required only in steps 2 and 3 in order to compute the value function $v_\theta$ entering the conditional choice probabilities P. This requires recomputing the fixed point $v_\theta$ of the contraction mapping (3.16) each time the outer nonlinear maximization algorithm computes new values for $\theta$. As discussed in section 3, if there are continuous state variables, then the fixed point $v_\theta$ 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 that i and the discretized values that x can assume. Suppose that $w_t$ is discretized into 100 cells, $y_t$ into 5 cells, and $c_t$ into 5 cells. Assuming that $aw_t$ can be proxied by $w_t$ and $y_t$, the remaining state variables assume the following number of values: $h_t$:4, $a_t$:20, $e_t$:3, $ms_t$:2, $d_t$: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 30 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 (3.16). 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 $v_0$:

(4.4) $\quad v_{k+1}=T(v_k)$ $\quad$ contraction iteration

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:

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

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 $\beta$ 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×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 special structure that can be exploited in order to dramatically reduce the computational burden of the fixed point problem. 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 $\pi$, 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 $v_\theta$ need only be calculated for the 3 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 worker's die with probability 1 beyond some fixed age (say 98), one can compute the fixed point $v_\theta$ 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 workers live beyond age 90. 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 $v_\theta$, increasing the required computer time by several orders of magnitude. Since the assumption of fixed lifespan is relatively innocuous and leads to substantial computational simplications, 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 $\pi$ and significantly reduce the number of operations needed to evaluate the contraction mapping T (3.16) or solve the N×N linear system in the Newton-Kantorovich iteration (4.5).

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 $\pi$. This is equivalent to left matrix multiplication of the "vectorized" integrand by the matrix representation of $\pi$.

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 $\pi$. 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 megflops 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 non-zero 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 3 through 8 depict different "sparsity patterns" for the matrix representation of $\pi$ depending on the ordering of the component state variables in $x_t$. $\pi$ can be regarded as a direct product of 3 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 3. By varying the order of these component matrices in the construction of the direct product, one obtains different sparsity patterns for $\pi$. Figures 4, 5, and 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 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 8 shows the packed storage arrangement for the block banded matrices that form the off-diagonal sectors of $\pi$. This arrangement allows one to fully exploit the sparsity of $\pi$ while keeping the vector processor running at nearly maximum efficiency.

(figures 3,4,5,6,7,8 here)

Exploitation of sparsity patterns is particularly important in the infinite horizon case. For sufficiently high discount factors $\beta$ 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 $\beta$ times the transition probability matrix for the controlled process $\{i_t,x_t\}$ which is isomorphic to the basic transition matrix for $\pi$. 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 $\pi$ in figures 4 through 7 except for the 1's along the diagonal. Under the "optimal" ordering (a,w,d), one can see from figure 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 speed-up of 8,000 times when a=20. Further speed-ups can be obtained by accounting for the block-banded structure of the (a,a) 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×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 occurs within the elementary d×d block operations of the block elimination procedure.[8] 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 fortiutous situation allows one to solve linear systems that are orders of magnitude larger than the largest systems solvable using standard sparse matrix software.

I conclude with table 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.[9] Overall table 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.

(table 6 here)

*This research is part of an NBER project on the economics of aging administered by David Wise, JFK School, Harvard University funded by National Institute for Aging grant 3 P01 AG05842-01. CPU time on the Cray-2 supercomputer was provided under grant SES-8419570 of the National Science Foundation.

[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" (footnote 7, page 405).

[3]I omit certain regularity conditions needed to insure existence of an optimal stationary policy. For details, see Rust (1986).

[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). However 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 (1986) 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 the having the simulation errors enter linearly and additively separably, directly extends to allow simulation instead of integration in the fixed point condition (3.8). In that case the simulation error is no longer additively separable, and the simulations must increase with the sample size to avoid inconsistency due to non-linear "errors in variables".

[8]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 the computed solution agreed with the theoretical solution to 12 significant digits.

[9]A simulation analysis of the Cray-2 processor (on a MacIntosh PC) by Lawrence Liddiard (1986) suggests that a single processor can achieve a maximum rate of 433 megaflops for dense matrix multiplication. In practice the highest rates that have been recorded for the University of Minnesota Cray-2 have been on the order of 360 megaflops. The single processor average of 220 megaflops reported in table 6 has been acheived using standard library kernels without special assembly language coding to optimize the flow of data from common memory to local memory and 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 $v_\theta$ and the remaining processors assigned to computing each of the derivatives $\partial v_\theta/\partial\theta$), allowing a sustainable processing rate of approaching 880 megflops.

Table 1: Distributions of Retirement Sequences[††]

| Sequence | Frequency | Sequence | Frequency |
|---|---|---|---|
| ffrr | 16.2% | frxx | 1.6% |
| ffff | 14.4 | rrrx | 1.5 |
| fffr | 11.2 | fppp | 1.4 |
| rrrr | 8.6 | frrx | 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 | fppr | 0.5 |
| ffrx | 2.5 | pqrr | 0.5 |
| rrxx | 2.2 | others | 9.8 |
| fprr | 2.1 | | |

Table 2: Labor Market Re-entry Rates

| | One year re-entry rates | | Two year re-entry rates | |
|---|---|---|---|---|
| Age | Self-described retired or un-able to work | Not full-time worker | Self-described retired or un-able uo 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 |

[††] Source: Gustman and Steinmeier (1986), page 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 indeterminant. Sequences with a frequency less than 0.5% were grouped in the category "others".

Table 3: Accounting for Labor Force Transitions in the Dynamic Programming Model

| | Employment state, $e_t$ | Employment decision, $d_t$ | Interpretation |
|---|---|---|---|
| 1. | ft | ft | continue working at current full-time job |
| 2. | ft | pt | quit current full-time job, search for a new part-time job |
| 3. | ft | ne | if $a_t \geq 62$, retire; if $a_t < 62$ and disabled, receive disability insurance; otherwise exit labor force |
| 4. | pt | ft | quit current part-time job and search for a full-time job |
| 5. | pt | pt | continue working at current part-time job |
| 6. | pt | ne | if $a_t \geq 62$, retire; if $a_t < 62$ and disabled, collect disability insurance; otherwise exit labor force |
| 7. | ne | ft | unemployed, disabled, or retired worker searching for full-time job |
| 8. | ne | pt | unemployed, disabled, or retired worker searching for part-time job |
| 9. | ne | ne | if $a_t \geq 62$, remain retired; if $a_t < 62$ and disabled, collect disability insurance; otherwise remain out of labor force |

Table 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 $i_t$ when the observed state variable is $x_t$. |
| $\varepsilon_t=\{\varepsilon_t(i)\mid i\in C(x_t)\}$ | 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. |
| $x_t=\{x_t(1),\ldots,x_t(M)\}$ | 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 $(x_t,\varepsilon_t)$ when alternative i is chosen. |
| $p(x_{t+1},\varepsilon_{t+1}\mid x_t,\varepsilon_t,i)$ | Markov transition density for next period state variable when alternative i is chosen and when the current state is $(x_t,\varepsilon_t)$. |

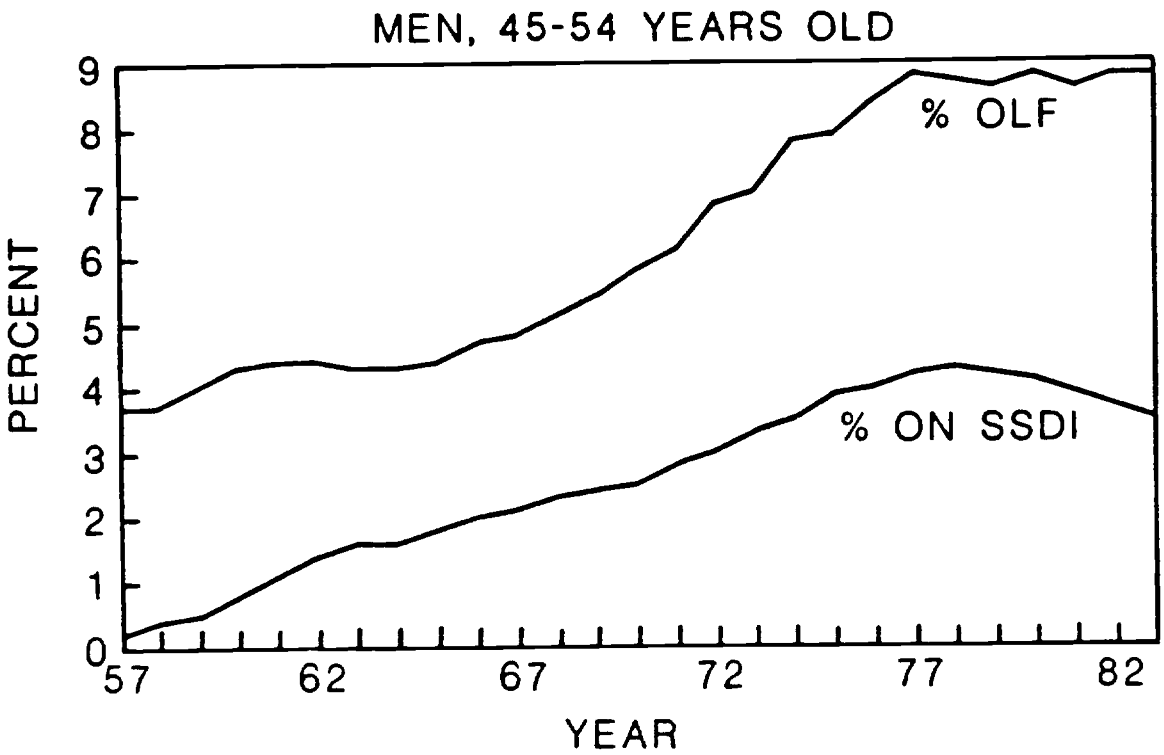
Table 5: Summary of the Retirement Decision Problem

| Item | Notation |
|---|---|
| 1. Choice set | $C(x)=\{1,2,3\}\otimes\{c^1,\ldots,c^J\}$, 1=ft, 2=pt, 3=ne |
| 2. Control vector | $i_t=(d_t,c_t)$; $d_t\in\{1,2,3\}$, $c_t\in\{c^1,\ldots,c^J\}$ |
| 3. State vector (observed) | $x_t=(w_t,h_t,a_t,ms_t,cs_t,e_t,y_t,aw_t)$ |
| 4. State vector (unobserved) | $\varepsilon_t\equiv\{\varepsilon_t(i)\mid i\in C(x_t)\}$, $\varepsilon_t(i)=\varepsilon_t(d,c)$ |
| 5. Bequest function | $b(w_t,ms_t,cs_t,\theta_1)$ |
| 6. Utility function | $u(d_t,c_t,e_t,h_t,a_t,ms_t,cs_t,\theta_2)$ |
| 7. Transition density ($x_t$) | $\pi(x_{t+1}\mid x_t,i_t,\theta_3)$ |
| 8. Transition density ($\varepsilon_t$) | $q(\varepsilon_t\mid x_t,\theta_4)\sim GEV(C(x_t),\theta_4)$, see (3.11) |
| 9. Parameter vector | $\theta=(\beta,\theta_1,\theta_2,\theta_3,\theta_4)$, $1\times(1+K_1+K_2+K_3+K_4)$ |

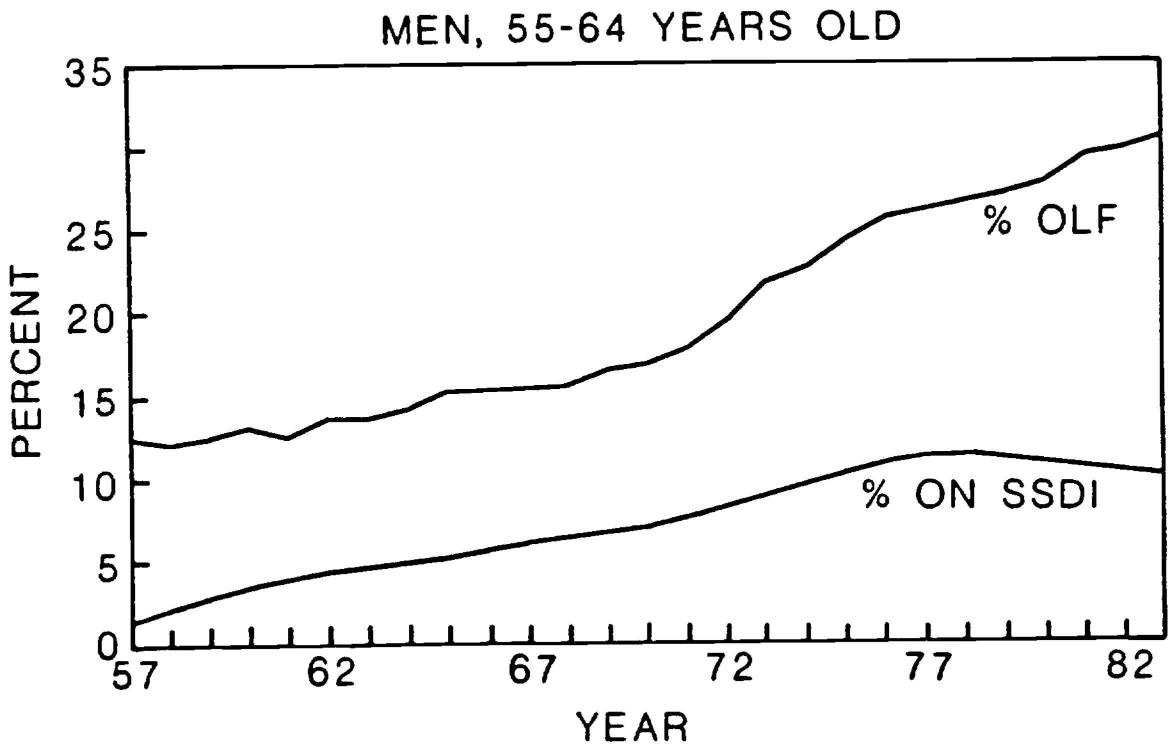
Table 6: Fixed Point Computation Times on the Cray-2[*]

| Item | | Finite Horizon | Infinite 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 |

[*] Times are for 1 processor on the University of Minnesota 4 processor Cray-2 with 256 million word common memory.



MEN, 45-54 YEARS OLD
% OLF
% ON SSDI
PERCENT
YEAR
0
1
2
3
4
5
6
7
8
9
57
62
67
72
77
82

MEN, 55-64 YEARS OLD
% OLF
% ON SSDI
PERCENT
YEAR
0
5
10
15
20
25
35
57
62
67
72
77
82


FIGURE 1

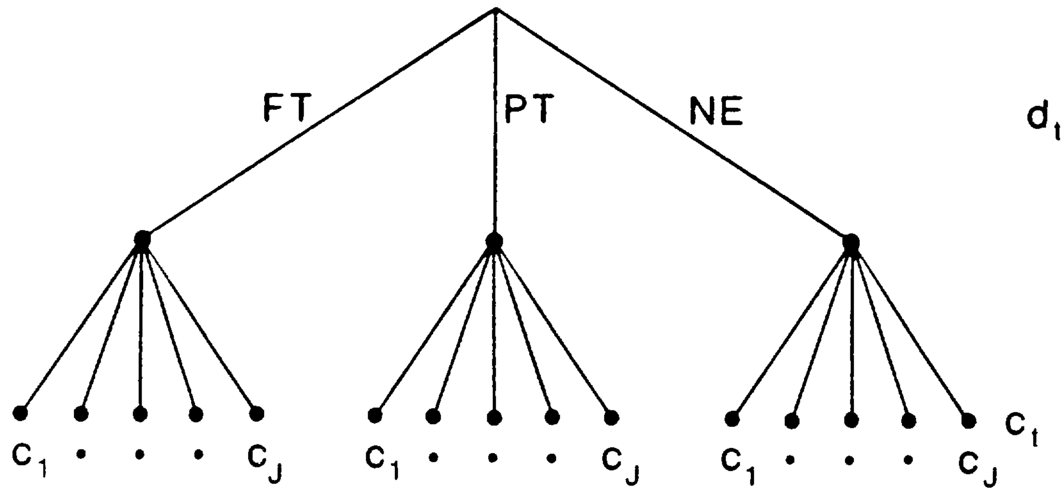



FIGURE 2

# State Transition Matrices

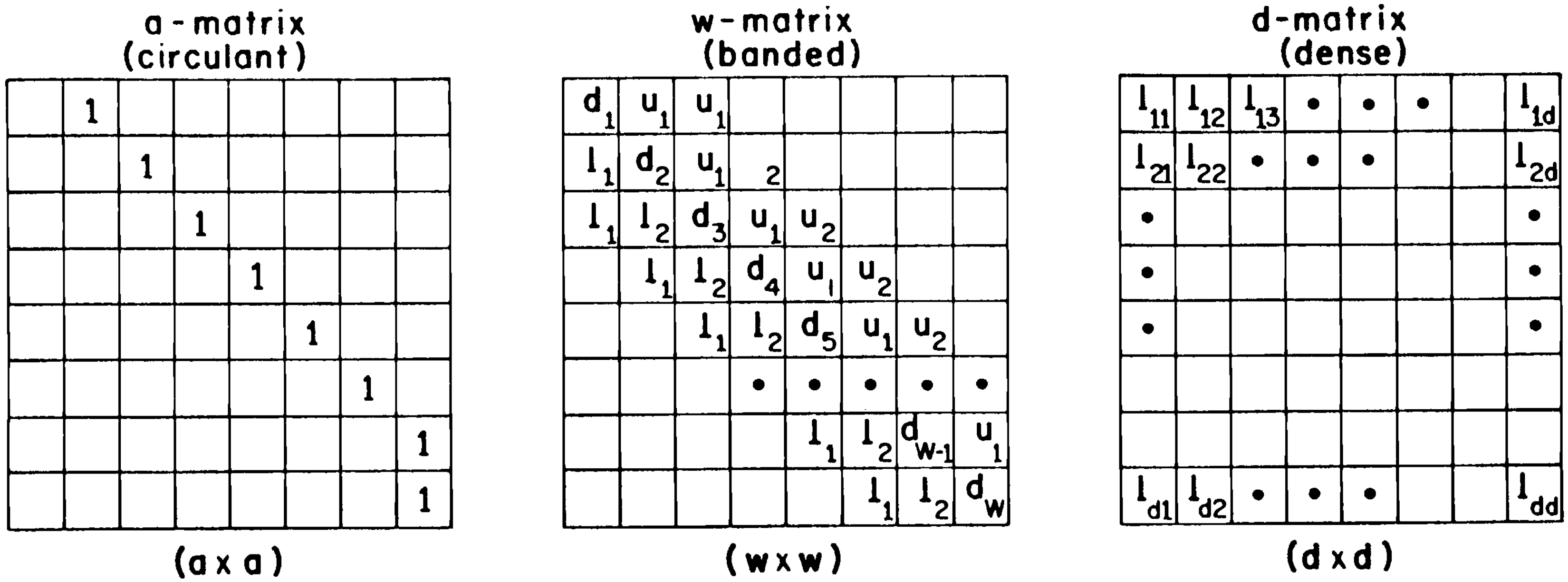
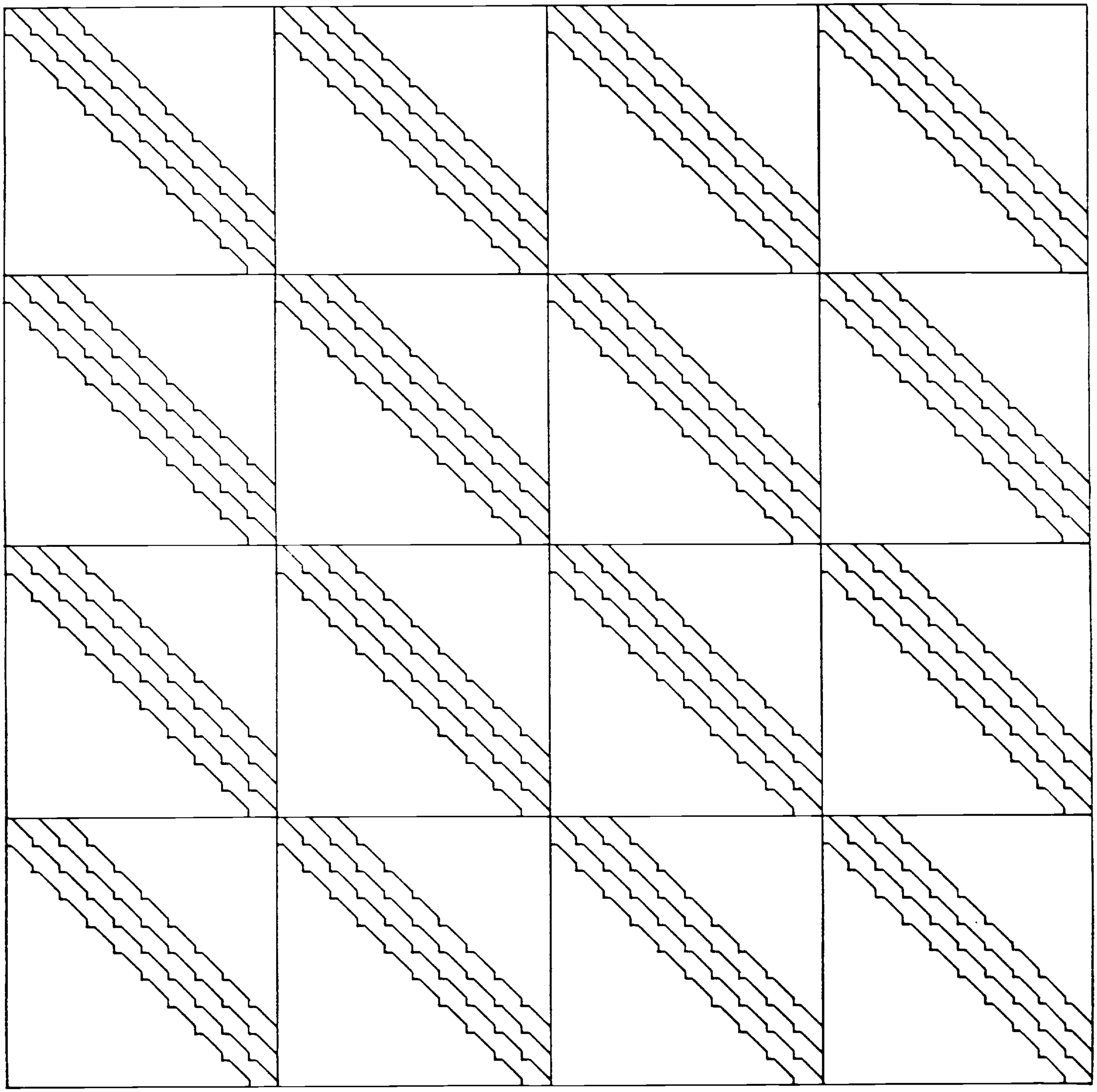


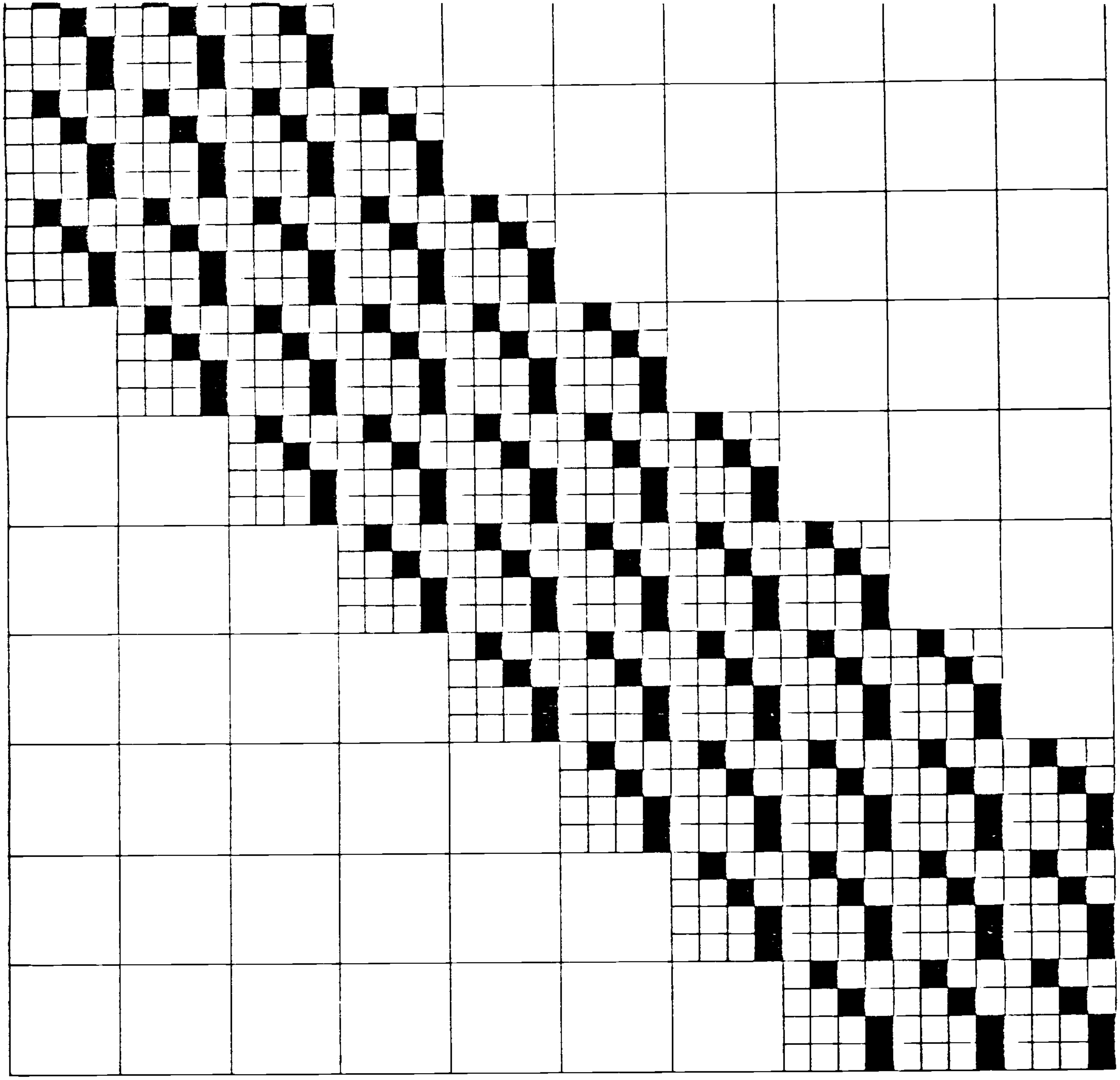

FIGURE 3



**SPARSITY PATTERN:**

**(d,w,a)**

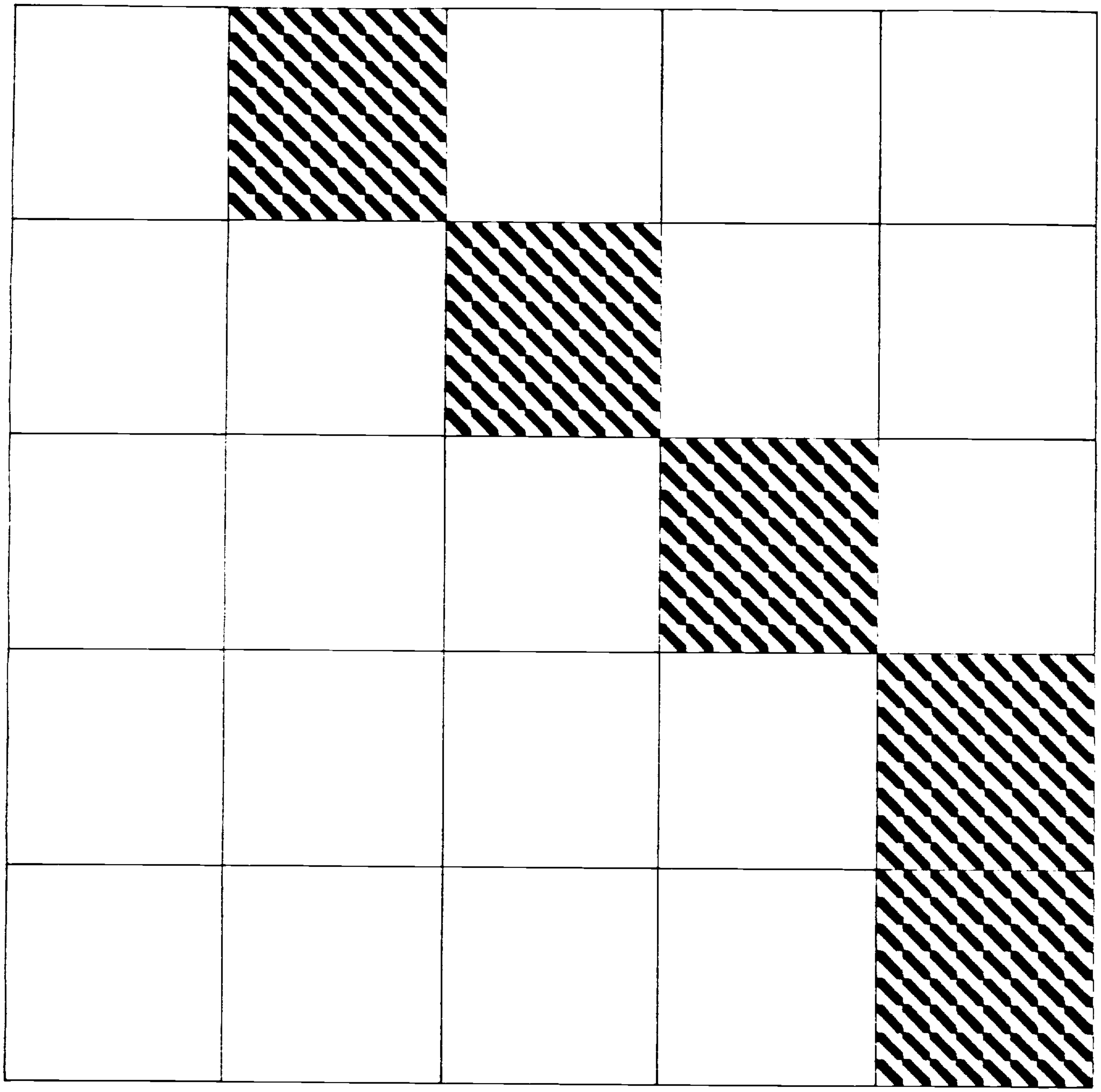
FIGURE 4



**SPARSITY PATTERN:**

**(w,a,d)**

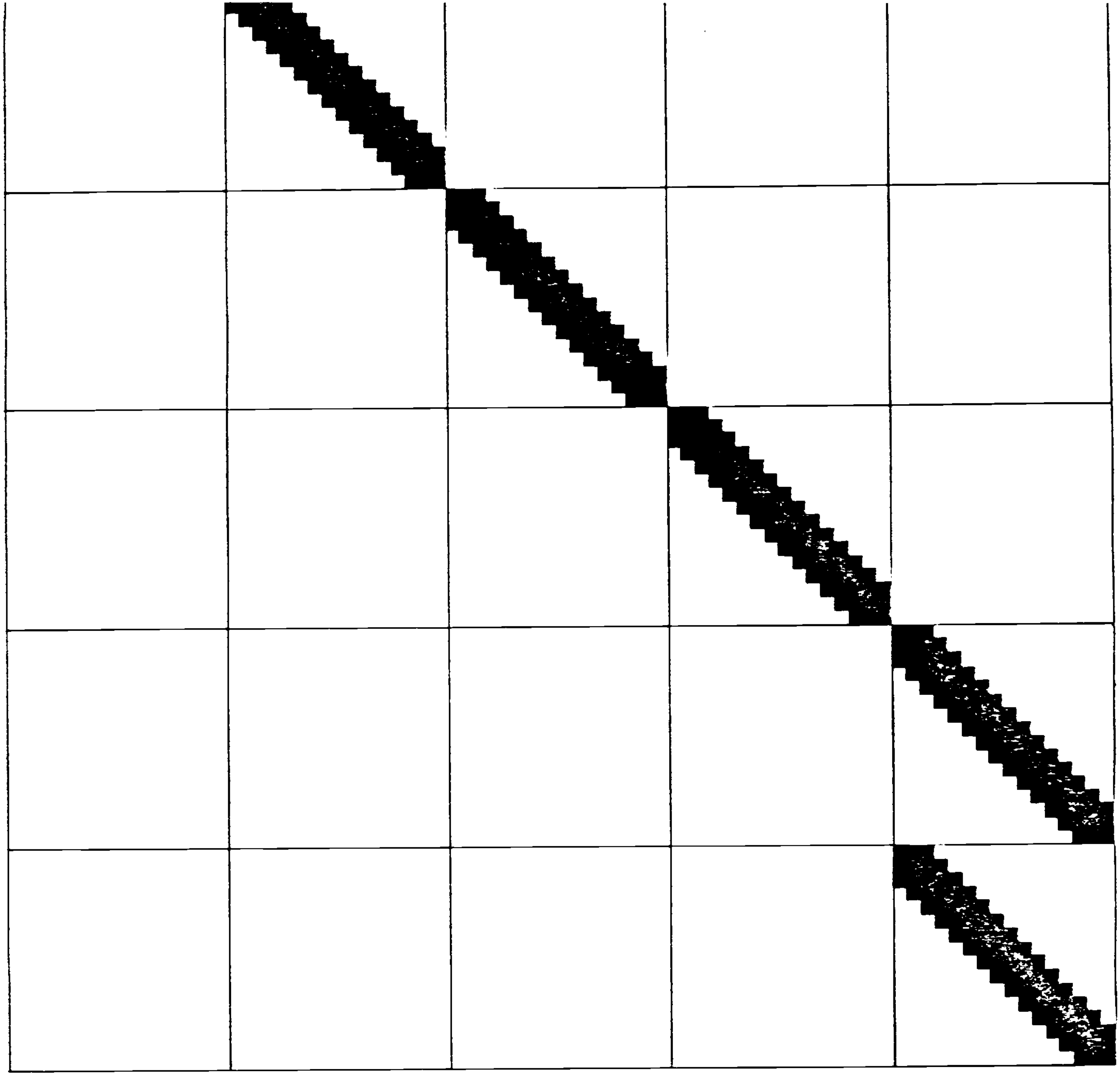
FIGURE 5



**SPARSITY PATTERN:**

**(a,d,w)**

FIGURE 6



**SPARSITY PATTERN:**

**(a,w,d)**

FIGURE 7

# PACKED STORAGE FORM FOR BLOCK-BANDED MATRIX

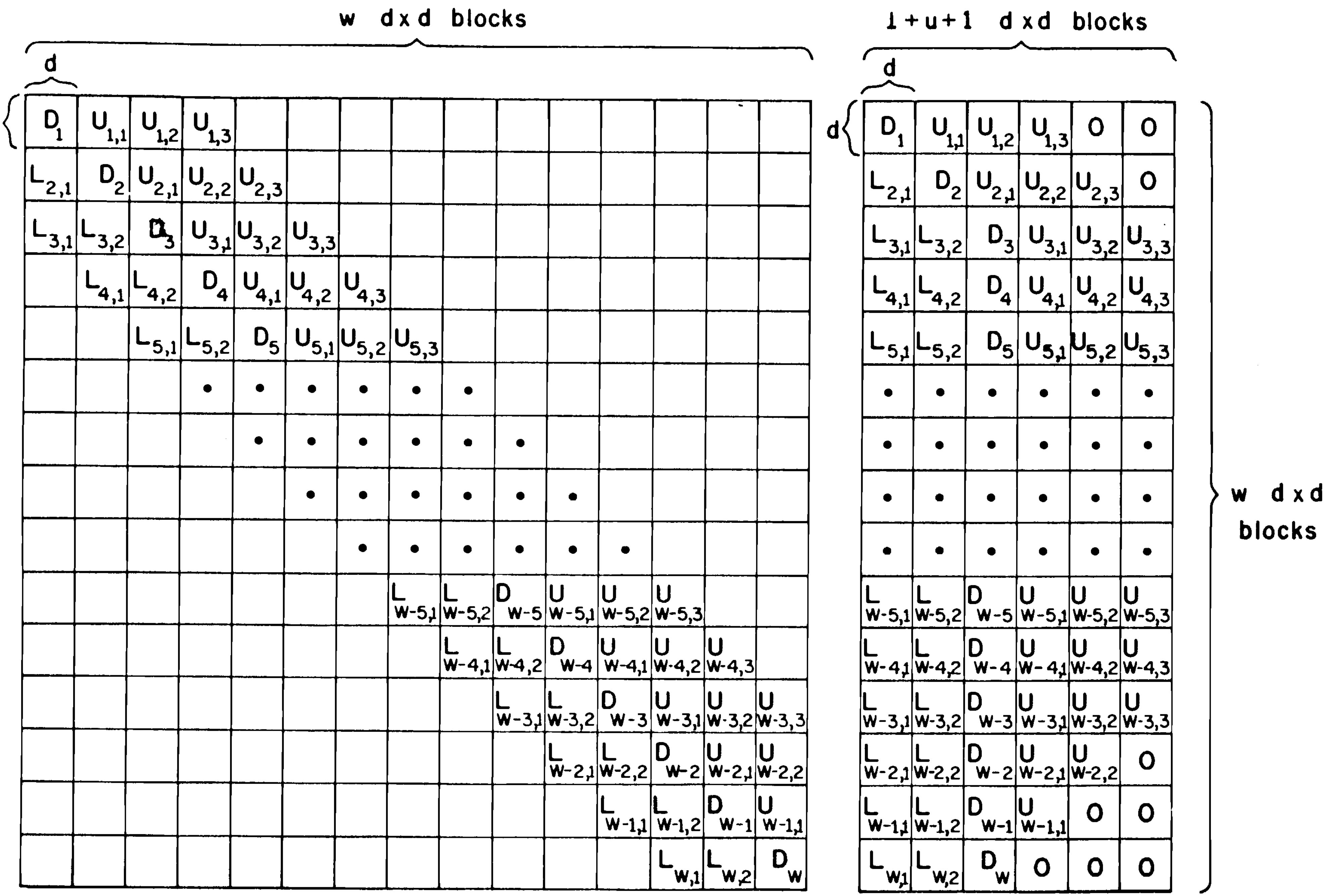



FIGURE 8

## Bibliography


Anderson, K.H. Burkhauser, R.V. Butler, J.S. (1984) "Work After Retirement: A Hazard Model of Labor Market Re-entry" manuscript.

Anderson, K.H. Burkhauser, R.V. Quinn, J.F. (1984) "Do Retirement Dreams Come True? The Effect of Unexpected Events on Retirement Age" DP #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.

Barro, R.J. (1978) *The Impact of Social Security on Private Saving* The American Enterprise Institute, Washington.

Berndt, E. Hall, B. Hall, R. Hausman, J. (1974) "Estimation and Inference in Nonlinear Structural Models" *Annals of Economic and Social Measurement* 3 653-665.

Bernheim, D. (1984) "Dissaving After Retirement: Testing the Pure Life Cycle Hypothesis" *NBER Working Paper* 1409.

Blackwell, D. (1965) "Discounted Dynamic Programming" *Annals of Statistics* 36, 226-235.

Blinder, A. Gordon, R.H. Wise, D.E. (1983) "Social Security, Bequests, and the Life-Cycle Theory of Savings" in F. Modigliani and R. Hemming (eds.) *The Determinants of National Saving and Wealth* New-York, St. Martin's Press, 89-122.

Bodie, Z. Shoven, J.B. (eds.) (1983) *Financial Aspects of the United States Pension System* University of Chicago Press.

Boskin, M. (1977) "Social Security and Retirement Decisions" *Economic Inquiry* January.

Boskin, M.J. Kotlikoff, L.J. (1985) "Public Debt and United States Savings: A New Test of the Neutrality Hypothesis" in K. Brunner and A.H. Meltzer (eds.) *The "New Monetary Economics", Fiscal Issues and Unemployment* Volume 23, North-Holland, Amsterdam.

Boskin, M.J. Kotlikoff, L.J. Puffert, D.J. Shoven, J.B. (1986) "Social Security: A Financial Appraisal Across and Within Generations" *NBER Working Paper 1891*.

Boskin, M., Hurd, M., (1978) "The Effect of Social Security on Early Retirement" *Journal of Public Economics* 10, 361-377.

Boskin, M.J. Shoven, J.B. (1986) "Poverty Among the Elderly: Where are the Holes in the Safety Net?" *NBER Working Paper* 1923.

Bound, J. (1986) "The Disincentive Effects of the Social Security Disability Program" manuscript, University of Michigan.

Burkhauser, R.V. (1980) "The Early Acceptance of Social Security: An Asset Maximization approach" *Industrial and Labor Relations Review* 33, 484-492.

Burkhauser, R.V. (1979) "The Pension Acceptance Decision of Older Workers" *Journal of Human Resources* 14-1, 63-75.

Burkhauser, R.V., Turner, J.A. (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, 701-715.

Burtless, G., Aaron, H. (eds) (1984) *Retirement and Economic Behavior* Brookings Institution, Studies in Social Economics.

Burtless, G. Moffitt, R.A. (1985) "The Just Choice of Retirement Date and Post-Retirement Hours of Work" *Journal of Labor Economics*, April.

Burtless, G., Moffitt, R. (1984) "The Effect of Social Security Benefits on the Labor Supply of the Aged" in Burtless, Aaron (eds) *Retirement and Economic Behavior*.

Cartwright, W.S. Friedland, R.B. (1985) "The President's Commission on Pension Policy Household Survey 1979" *Review of Income and Wealth* 31, 285-308.

Chirikos, T.N. Nestel, G. (1981) "Impairment and Labor Market Outcomes: A Cross-Sectional and Longitudinal Analysis" in Parnes, H. (ed.) *Work and Retirement*, MIT Press, Cambridge.

Congressional Quarterly Inc. (1983) *Social Security and Retirement* Congressional Quarterly, Washington.

Crawford, V.P., Lilien, D.M. (1981) "Social Security and the Retirement Decision" *Quarterly Journal of Economics* August, 505-529.

Danziger, S., Haveman, R. Plotnick, R. (1981) "How Income Transfers Affect Work, Saving, and the Income Distribution" *Journal of Economic Literature* 19-3, 975-1028.

Danziger, S. van der Gaag, J. Smolensky, E. Taussig, M. (1982) "The Life-Cycle Hypothesis and the Consumption Behavior of the Elderly" *Journal of Post-Keynesian Economics* 5, 208-227.

Darby, M.R. (1979) *The Effects of Social Security on Income and the Capital Stock* American Enterprise Institute, Washington, D.C.

Davies, J.B. (1981) "Uncertain Lifetime, Consumption, and Dissaving in Retirement" *Journal of Political Economy* 89-3, 561-577.

Deaton, A. (1986) "Life-Cycle Models of Consumption: Is the Evidence Consistent with the Theory?" NBER Working Paper 1910.

Diamond, P. (1977) "A Framework for Social Security Analysis" *Journal of Public* Economics 8-3, 275-298.

Diamond, P. Hausman, J. (1984a) "Individual Retirement and Savings Behavior" *Journal of Public Economics* 23, 81-114.

Diamond, P.A., Hausman, J.A. (1984b) "The Retirement and Unemployment Behavior of Older Men" in Burtless, Aaron (eds) *Retirement and Economic Behavior.*

Diamond, P. Mirrlees, J.A. (1978) "A Model of Social Insurance with Variable Retirement" *Journal of Public Economics* 10, 295-336.

Dubin, J. McFadden, D. (1984) "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption" *Econometrica* 52-2, 345-362.

Duff, I.S. (1984) "The Solution of Sparse Linear Equations on the Cray-1" in J.S. Kowalik (ed) *High Speed Computation* Springer-Verlag, Berlin.

Feldstein, M. (1974) "Social Security, Induced Retirement and Aggregate Capital Accumulation" *Journal of Political Economy* 82, 905-926.

Feldstein, M. (1983) "Social Security Benefits and the Accumulation of Pre-retirement Wealth" in F. Modigliani and R. Hemming (eds.) *The Determinants of National Wealth and Saving* St. Martin's Press, New York 3-23.

Felstein, M. Pellechio, A.J. (1979) "Social Security and Household Wealth Accumulation: New Microeconometric Evidence" *Review of Economics and Statistics* 61-3, 361-368.

Feldstein, M.S. Pellechio, A. (1979) "Social Security and Household Wealth Accumulation: New Microeconomic Evidence" *Review of Economics and Statistics* 61, 361-368.

Fields, G.S. Mitchell, O.S. (1985) *Retirement, Pensions, and Social Security* M.I.T. Press.

Fuchs, V.R. (1982) "Time Preference and Health: An Exploratory Study" in V. Fuchs (ed.) *Economic Aspects of Health* University of Chicago Press.

Gordon, R.H. Blinder, A.S. (1980) "Market Wages, Reservation Wages, and Retirement Decisions" *Journal of Public Economics* 14, 277-308.

Gotz, G.A. McCall, J.J. (1984) "A Dynamic Retention Model for Air Force Officers" Research report R-3028-AF, The Rand Corporation, Santa Monica, California.

Gustman, A.L. Steinmeier, T.L. (1983) "Minimum Hours Constraints and Retirement Behavior" *Economic Inquiry* 3, 77-91.

Gustman, A.L. Steinmeier, T.L. (1984) "Partial Retirement and the Analysis of Retirement Behavior" *Industrial and Labor Relations Review* 37, 403-415.

Gustman, A. Steinmeier, T.L. (1986) "A Structural Retirement Model" *Econometrica* 54-3, 555-584.

Hall, A. Johnson, T. (1980) "The Determinants of Planned Retirement Age" *Industiral and Labor Relations Review* 33-2, 241-255.

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, 541-562.

Hansen, L.P. Singleton, K. (1982) "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models" *Econometrica* 50-5, 1269-1286.

Harris, L. (1984) *Retirement and Income* Garland Publishing, New York.

Hurd, M.D. (1986) "Savings and Bequests" *NBER Working Paper Series* 1826.

Irelan,L. Motley, D. Schwab,K. Sherman, S. Murray, J. (1976) *Almost 65: Baseline Data for the Retirement History Study* HEW Research Report 49.

King, M. (1985) "The Economics of Saving: A Survey of Recent Contributions" in K. Arrow and S. Harkapohja (eds.) *Frontiers of Economics* Basil Blackwell, Oxford.

King, M. Dicks-Mireaux, L.D. (1982) "Asset Holdings and the Life-Cycle" *The Economic Journal* 92, 247-267.

Kotilkoff, L.J. (1979) "Testing the Theory of Social Security and Life Cycle Accumulation" *American Economic Review* 69, 396-410.

Kotlikoff, L.J. (1986) "Health Expenditures and Precautionary Savings" *NBER Working Paper Series* 2008.

Kotlikoff, L.J. Smith, D.E. (1983) *Pensions in the American Economy* University of Chicago Press.

Kotlikoff, L.J. Summers, L.H. (1981) "The Role of Intergenerational Transfers in Aggregate Capital Accumulation" *Journal of Political Economy* 89, 706-732.

Kotlikoff, L.J. Spivak,A. Summers, L.H. (1982) "The Adequacy of Savings" *American Economic Review* 72-5, 1056-1069.

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.

Kurz, M. (1985) "Hetrogeneity and Savings Behavior: A Comment" in K. Arrow and S. Harkapohja (eds.) *Frontiers of Economics* Basil Blackwell, Oxford.

Kurz, M. (1984) "Capital Accumulation and the Characteristics of Private Intergenerational Transfers" *Economica* 51, 1-22.

Liddiard, L.D. (1986) "How to get the most out of a Cray-2" manuscript, University of Minnesota.

Lucas, R.E. (1976) "Econometric Policy Evaluation: A Critique" in Brunner, K. Meltzer,A.K. (eds) *The Philips Curve and Labor Markets Carnegie-Rochester* **Conference on Public Policy 1**, North Holland.

MacCurdy, T.E. (1983) "A Simple Scheme for Estimating an Intertemporal Model of

Sectional View of White Married Men"

Labor Supply and Consumption in the Presence of Taxes and Uncertainty" *International Economic Review* 24-2, 265-289.

Marschak, J. (1953) "Economic Measurements for Policy and Prediction" in Hood, W.C., Koopmans, T.C. (eds) *Studies in Econometric Method* Wiley.

McFadden, D. (1987) "A Method of Simulated Moments for Estimation of Multinomial Probits without Numerical Integration" *Econometrica*, forthcoming.

Menchik, P.L. David, M. (1983) "Income Distribution, Lifetime Savings, and Bequests" *American Economic Review* 73-4, 672-690.

Menchik, P.L. David, M. (1985) "The Effect of Social Security on Lifetime Wealth Accumulation and Bequests" *Economica* 52-208, 421-434.

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, 435-443.

Mirer, T. (1980) "The Dissaving Behavior of the Aged" *Southern Economic Journal* 46, 1197-1205.

Mitchell, O.S., Fields G.S. (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 M. Pakin (ed.) *Contemporary Issues in Economics* Manchester University Press.

Modigliani, F. Brumberg, F. (1954) "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data" in K. Kurihara (ed.) *Post-Keynesian Economics* New Brunswick.

Morrison, M.H. (1982) *Economics of Aging: The Future of Retirement* van Nostrand Rheinhold, New York.

Mott, F.L. Haurin, R.J. (1981) "The Impact of Health Problems and Mortality on Family Well-Being" in Parnes, H. (ed.) *Work and Retirement*, MIT Press, Cambridge.

Munnell, A.H. (1977) *The Future of Social Security* Brookings Institution, Washington, D.C.

Pakes, A. (1986) "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks" *Econometrica* 54-4, 755-784.

Pakes, A. Pollard, D. (1986) "The Asymptotics of Simulation Estimators" manuscript, University of Wisconsin.

Parnes, H.S. (editor) (1981) *Work and Retirement* MIT Press, Cambridge.

Parnes, H.S. (editor) (1983) *Policy Issues in Work and Retirement* W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.

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, 284-287.

Quinn, J. F. (1977) "Microeconomic Determinants of Early Retirement: A Cross-Sectional View of White Married Men" *Journal of Human Resources* 12, 329-346.

Quinn, J.F. Burkhauser, R.V. (1983) "Is Mandatory Retirement Overrated?: Evidence from the 1970's" *Journal of Human Resources* 18-3, 337-358.

Rust, J. (1986) "Maximium Likelihood Estimation of Discrete Control Processes" SSRI Working Paper 8407.

Rust, J. (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher" *Econometrica*, forthcoming.

Rust, J. (1987a) "An Inverse Solution Algorithm for Structural Estimation of Discrete Control Processes"

Sargent, T.J. (1981) "Interpreting Economic Time Series" *Journal of Political Economy* 89-2, 213-248.

Schulz, J.H.(1976) *The Economics of Aging* Wadsworth Publishing, Belmont, California.

Sickles, R.C. Taubman, P. (1986) "An Analysis of the Health and Retirement Status of the Elderly" *Econometrica* 54-6, 1339-1356.

Sheshinski, E. (1978) "A Model of Social Security and Retirement Decisions" *Journal of Public Economics* 10, 337-360.

Stein, B. (1980) *Social Security and Prensions in Transition* Free Press, Macmillan Publishing Co., New York.

Sueyoshi, G. (1986) "Social Security and the Determinants of Full and Partial Retirement: A Competing Risks Analysis" manuscript, M.I.T.

Thompson, L.H. (1981) "The Social Security Reform Debate" *Journal of Economic Literature* 21-4, 1425-1467.

Tobin, J. (1967) "Life-Cycle Saving and Balanced Growth" in W. Fellner (ed.) *Ten Economic Studies in the Tradition of Irving Fisher* New-York, 231-256.

White, B. (1978) "Empirical Tests of the Life-Cycle Hypothesis" *American Economic Review* 68, 547-560.

White, B. (1984) "Empirical Tests of the Life-Cycle Hypothesis: Reply" *American Economic Review* 74, 258-259.

Williams, C.A.jr. Turnbull, J.G. Cheit, E.F. (1982) *Economic and Social Security: Social Insurance and Other Approaches* Wiley, New York.

Wise, D. (ed.) (1985) *Pensions, Labor and Individual Choice* University of Chicago Press.

Wolpin, K. (1984) "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality" *Journal of Political Economy* 92-5, 852-874.

Yaari, M.E. (1965) "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer" *Review of Economic Studies* 32, 137-150.

Levhari, D. Mirman, L.J. (1977) "Savings and Consumption with an Uncertain Horizon" *Journal of Political Economy* 85-2, 265-281.

Zabalza, A., Pissarides, C., Barton, M. (1980) "Social Security and the Choice Between Full-time Work, Part-time Work, and Retirement" *Journal of Public Economics* 14, 245-276.