Econometrica, Vol. 59, No. 1 (January, 1991), 189–210

JOB EXIT BEHAVIOR OF OLDER MEN

By James Berkovec and Steven Stern[1]

We estimate a dynamic programming model of job exit behavior and retirement using the method of simulated moments. The model and estimation method allow for both unobserved individual effects and unobserved job-specific "match" effects. The model is estimated using two different assumptions about individual discount factors. First, a static model, with the discount factor β equal to zero, is estimated. Then a dynamic model, with $\beta = .95$, is estimated. In both models, it is found that bad health, age, and lack of education increase the probability of retirement. The dynamic model performs better than the static model and has different implications for retirement behavior. The job-specific effects are an important source of unobserved heterogeneity.

Keywords: Dynamic programming model, retirement, job switching, method of simulated moments, unobserved heterogeneity, match effects.

1. INTRODUCTION

Labor force participation decisions such as retirement have been a major research topic for economists. Much of the work of retirement and job exit behavior has included dynamics but in somewhat restricted ways. For example, Heckman and MaCurdy (1980) and MaCurdy (1981) control for life cycle effects by adding a fixed effect that is a function of the marginal utility of wealth at the beginning of life. This requires them to assume that the wage schedule facing an individual is independent of the individual's participation decisions, an assumption that is inconsistent with a job tenure effect on wages or with job-specific "match" effects.

This paper develops a model of job exit behavior and retirement for older men. The model assumes that workers choose the job state (full-time work, part-time work, or retirement) that has the highest value. The value of each state is defined by the contemporaneous flows, both pecuniary (wage) and nonpecuniary (leisure), to that state as well as the value of future flows conditioned on the past and current job state choices. The flows to each state include both unobservable individual and unobservable job-specific "match" effects as well as deterministic functions of observed characteristics and parameters. The value of the future flows conditioned on past and current choices is calculated naturally as the solution to a dynamic programming problem.

The model is formulated as a discrete-time, discrete-state, dynamic programming problem. Each individual chooses when to make job state changes and which changes to make at any time. When working, an individual chooses between keeping his existing job, accepting a new part-time or new full-time job, or retiring (not working). When retired, his choices are to remain retired or to

[1] We would like to thank Alan Gustman, Tom Steinmeier, Bill Johnson, John Rust, Dan McFadden, Ariel Pakes, and workshop participants at NBER, Virginia, Penn, Duke, Yale, Kentucky, Brown, and the Econometric Society for valuable comments, Shannon Mitchell and Diane Lim for careful research assistance, and Betty Smith for excellent typing. All remaining errors are ours. This work was funded by the National Institute on Aging, Grant No. 5R29AG06387.

accept either a new full-time or a new part-time job. The model developed in this paper allows workers to have only very limited uncertainty about the future. The model can be extended easily to include more realistic information about future options. Such an extension of the model, however, will expand the state space of the dynamic programming problem and consequently increase the computational requirements of the estimation procedure.

The estimation procedure used in this paper is a simulation estimator based on Pakes and Pollard (1989) and McFadden (1989). Unobserved individual effects are simulated using a Monte Carlo procedure and are used to form unbiased estimates of the computationally intractable predicted values from the model. These simulated values are then used to form moment conditions that are solved to find parameter estimates. The method of simulated moments provides a computationally feasible alternative to intractable maximum likelihood or method of moments estimation which require numerical integration of the unobserved individual and job-specific effects.

The estimation of dynamic programming models, like that of this paper, has been receiving considerable attention recently. Notable papers include the job matching model of Miller (1984), the fertility model of Wolpin (1984), the patent renewal problem of Pakes (1986), the labor force participation models of Eckstein and Wolpin (1986) and Gustman and Steinmeier (1986), and the durable replacement and retirement models of Rust (1984, 1987, 1988).[2] These papers all have the common feature of estimating a dynamic discrete choice model, but differ substantially in their stochastic structures and in their estimation techniques. The approach used in this paper is distinguished by the flexible correlation structure of unobservables, the inclusion of unobservable individual and job-specific effects,[3] the use of continuous dependent variables (wages) in addition to the discrete events, and the use of the method of simulated moments for estimation.

The next section of the paper discusses the economic model and develops the dynamic programming problem. The model developed here is a partially reduced form in that, for example, the specific features of the Social Security system are omitted. A more structural model that includes a detailed representation of the structure of Social Security benefits is feasible in this framework but is left for future work. Section 3 discusses the data, a subset of the National Longitudinal Survey of Mature Men, used in estimation. Sections 4 and 5 discuss the estimation procedure and results respectively. Section 6 concludes.

2. THE DYNAMIC PROGRAMMING PROBLEM

Each individual's job history is modelled as a finite horizon, discrete time, dynamic programming problem with periods $t = 1, 2, \dots, T^*$. Each individual makes a job change/retirement decision at the beginning of each period based

[2] See Eckstein and Wolpin (1989) for a survey of empirical papers using a dynamic programming approach.

[3] Miller (1984) also allows for job-specific effects.

on observable values of pecuniary (wages and pensions) and nonpecuniary (leisure) flows as well as expectations about future pecuniary and nonpecuniary flows that are contingent on current and past job decisions. Full-time and part-time job offers are assumed to arrive each at a rate of one per period with the utility flow from each opportunity containing a known, job-specific error. The utility flow from each opportunity also contains a time-specific error that is known for the current period but unknown for future periods.

Define $V(F,t,s)$ as the value at time t of being employed at a full-time job that started at time s. Similarly, define $V(P,t,s)$ as the value at time t of being employed at a part-time job that started at s and $V(R,t,s)$ as the value at t of being retired since time s. An individual working at time t has four options at $t+1$: (a) continue with the same job, (b) switch to a new full-time job, (c) switch to a new part-time job, or (d) retire. An individual retired at time t has three options at $t+1$: (a) remain retired, (b) accept a new full-time job, or (c) accept a new part-time job. The individual chooses the option corresponding to the maximum value.

The values of being employed full-time, being employed part-time, and being retired are specified, respectively, as

$$\begin{aligned}(2.1)\qquad V(F,t,s) = {} & \bar{w}_F(t,s,X) + \tilde{w}_F(t,s,X) - cI(t,s)\\ & + \varepsilon_F(t,s) + \bar{\mu}_F(s) + \tilde{\mu}_F(s) + \beta_t EZ_F(t,s,X),\end{aligned}$$

$$\begin{aligned}(2.2)\qquad V(P,t,s) = {} & \bar{w}_P(t,s,X) + \tilde{w}_P(t,s,X) - cI(t,s)\\ & + \varepsilon_P(t,s) + \bar{\mu}_P(s) + \tilde{\mu}_P(s) + \beta_t EZ_P(t,s,X),\end{aligned}$$

and

$$\begin{aligned}(2.3)\qquad V(R,t,s) = {} & \bar{w}_R(t,s,X) + \tilde{w}_R(t,s,X)\\ & + \varepsilon_R(t,s) + \bar{\mu}_R + \tilde{\mu}_R + \beta_t EZ_R(t,s,X).\end{aligned}$$

The first right-hand terms, $\bar{w}_i$, $i = F, P, R$, are the deterministic part of the wage,[4] which depend upon age t, tenure $t-s$, and other exogenous variables $X(t)$. The next terms, $\tilde{w}_i$, $i = F, P, R$, are the deterministic part of the nonwage flows to utility, which depend upon the same arguments. The full-time and part-time equations have a cost of moving term c that is incurred only when starting a new job. The indicator function $I(t,s)$ is equal to one if $t=s$ and zero if $t>s$. The errors $\varepsilon_i(t,s)$ are the random components of the wage which are independent across time, matches, and individuals. The errors, $\bar{\mu}_F(s)$ and $\bar{\mu}_P(s)$, are match-specific random components of the wage, independent across matches and individuals but constant over the tenure of the job started at s. The error $\bar{\mu}_R$ is the retirement-specific random component of the retirement wage. The errors, $\tilde{\mu}_F(s)$, $\tilde{\mu}_P(s)$, and $\tilde{\mu}_R$, are the random components to nonwage utility flows. An individual knows the current values of ε_i and all current and future values of $\bar{\mu}_i$ and $\tilde{\mu}_i$ for $i = F, P, R$. Only the future values of ε_i are unknown by the individual. The discount factor, $\beta_t = \beta\delta_t$, includes a constant

[4] The retirement wage is any Social Security or pension benefit.

time preference component β (unobserved by the econometrician) and a time-varying mortality component δ_t (observed by the econometrician).[5] The last terms $EZ_i(t,s)$ are the expected values of the individual's best option at $t+1$ given information available at t:

$$(2.4) \quad EZ_F(t,s) = E\max[V(F,t+1,s), V(F,t+1,t+1), V(P,t+1,t+1), V(R,t+1,t+1)],$$

$$(2.5) \quad EZ_P(t,s) = E\max[V(P,t+1,s), V(F,t+1,t+1), V(P,t+1,t+1), V(R,t+1,t+1)],$$

and

$$(2.6) \quad EZ_R(t,s) = E\max[V(R,t+1,s), V(F,t+1,t+1), V(P,t+1,t+1)].$$

In general, only numerical solutions for $EZ_i(t,s)$ are available. If, however, the ε's have independent extreme value distributions:

$$(2.7) \quad F(\varepsilon_i) = \exp\{-e^{-\varepsilon_i/\tau}\}, \quad i = F, P, R,$$

then analytic solutions exist:

$$(2.8) \quad E\max_i[\bar{V}_i + \varepsilon_i] = \tau\left[\gamma + \ln\left(\sum_i e^{\bar{V}_i/\tau}\right)\right]$$

where $\bar{V}_i = V_i - \varepsilon_i$ and γ is Euler's constant.[6]

These value functions arise from a lifetime, time-separable utility maximization problem where consumption and leisure are separable and the individual's leisure choice is discrete. The effect of wealth on leisure enters through the $\tilde{\mu}$ terms because they can be functions of unobserved wealth at time $t = 1$. Essentially, the individual can decompose his maximization problem into two parts: first maximize V, and then allocate consumption across periods given V and unobserved initial wealth.

The dynamic programming problem can be solved backwards recursively. Assume there is some time T^* by which all individuals have retired permanently and a time T^{**} by which all individuals have died. Then the solution of the dynamic programming problem at T^* is trivial because there are no choices for the individual to make. The value of T^* should be large enough so that $\prod_{t=1}^{T^*}\beta_t$ is small (decisions made after T^* have small effects on early decisions) or so that no one chooses to work after T^* anyway.[7] Once the value functions are solved at T^*, they can be solved backwards recursively for all $t < T^*$ using equations (2.1), (2.2), (2.3), and (2.8).

[5] The mortality components are taken from Census life tables and control for age, cohort, and race. These represent an approximation to the actual expectations of future mortality used by individuals.

[6] See, for example, Dubin and McFadden (1984).

[7] In the empirical work in this paper, $T^* = 76$ and $T^{**} = 86$.

3. DATA

The data used in this study come from the National Longitudinal Survey of Mature Men (NLS). The NLS data set contains observations on 5020 men between the ages of 45 and 59 in 1966. These individuals are followed from 1966 until 1983 with interviews conducted approximately every other year. From the original NLS data, a sample of 2497 observations with complete job histories was constructed. A complete job history contains information on job status and individual characteristics for each year the individual was in the sample. Job status was imputed in nonsurvey years from information collected in adjacent survey years. Individuals without complete job histories (i.e., those who had too much missing data to reasonably impute a job state during each year of inclusion in the survey) were dropped from the sample. Each completed job history contains up to 17 years of information. Individuals who left the sample prior to 1983 but have complete job histories up until their exit are included using the available truncated information.

The job histories record job state in one of three categories, full-time work, part-time work, and not working, for each year. Job tenure is defined as the number of years since the individual started his current job. All changes of hours which result in a switch from full-time to part-time status or vice versa are considered to be changes in jobs which result in tenure clocks being reset. The job histories thus do not distinguish a significant shift of hours on the current job from a switch to a different job. In future efforts, it may be possible to separate these cases.

The health variable is constructed from answers to "limits" questions of the forms "Does your health limit the kind or amount of work you can do?" and "Does your health prevent you from working?" "Limits" questions may have subjective answers, but they are the only health related questions consistently asked in all eleven interviews.[8] Health is coded as 0 if the individual is healthy, 2 if unhealthy, and 1 if health status is uncertain. In noninterview years, if the individual had the same health status in both adjacent surveys he was assigned the same status. Health status is coded as uncertain for noninterview years where the adjacent interviews are different or for interview years when conflicting health data was obtained. Health data also needs to be simulated for the post survey period.[9] The procedure used to do this is to assume that each individual has a fixed probability of entering bad health in any post survey year, and that, once bad health occurs, it continues for the rest of the individual's life.

The current model assumes that an individual knows his health status for the rest of his life. The model can be expanded to include uncertainty about future health by including health status as another state variable. This approach was not used in this analysis in order to reduce computational costs.

[8] Stern (1989) suggests that a "limits" question does as well as any other health status measure in predicting labor force participation.

[9] Computation of value functions requires assumptions about health status up to time T^{**}.

TABLE I

FREQUENCIES OF OBSERVED JOB SPELL SEQUENCES

A		B	
Job Spell Sequence	Number of Observations	Job Spell Sequence	Number of Observations
fR	999	*fR*	206
f	509	*R*	73
R	187	*f*	51
fRF	122	*fRF*	26
fP	79	*fPR*	24
fPR	78	*fP*	19
fRFR	47	*pR*	13
fFR	42	*fPF*	9
fPF	37	*fPFR*	7
fPFR	35	*fFR*	7
pR	29	*fRFR*	7
fRP	16	*fF*	7
p	13	Others	51
fFF	13		
fRPR	12		
RFR	12		
fFPR	11		
fFP	11		
fFFR	11		
fPFP	10		
fRFFR	10		
Others	214		

Notes: (1) f is full-time job held at start; p is part-time job held at start; F is new full-time job; P is new part-time job; R is retirement. (2) Each letter represents a spell occurring over one or more years. (3) Observed sequences end either at the end of the sample period or when the individual is dropped from the sample due to death or nonresponse. The last spell observed may be incomplete. (4) Sample A is the full complete job history sample of 2497. Sample B is the estimation subsample of 500.

The characteristics of the data are described in Tables I through V. Two data files are described in Table I: the sample of 2497 with complete job histories and a subsample of 500 that was used in estimation. The estimation subsample is limited to individuals 55 and older at the start of the survey. The age limitation and the reduction in sample size[10] were both done solely to reduce the computational requirements of estimation.[11]

As shown in Table I, the most common job pattern, followed by 40% of the sample, is a single, full-time job, followed by retirement. Another 30% of the

[10] More than 500 observations meet the age restrictions. The 500 observations used were randomly sampled from those that were 55 and older. The estimation sample displays no significant, unexplainable differences from the larger sample.

[11] The major computational burden in estimation is the solution of the dynamic programming problem which must be done for each observation for every trial value of the parameters. The computation required to solve each individual's DP problem is proportional to the number of $\bar{V}$ values that must be computed. For this model, $[3 \cdot T^*(T^*+1)/2] + T^*$ future values are required where T^* is the time horizon. These values correspond to all states (combinations of age, job type, and tenure) that are attainable by the individual.

TABLE II

JOB STATE BY AGE OF INDIVIDUAL
ESTIMATION SAMPLE

Age	Retirement	New Full-Time	New Part-Time	Old Full-Time	Old Part-Time
55–59	298	34	16	1,090	39
60–64	740	52	29	1,036	94
65–69	991	12	40	329	167
70 +	568	12	14	128	112

TABLE III

JOB TRANSITION COUNTS
ESTIMATION SAMPLE

From To	Retirement	New Full-Time	New Part-Time	Old Full	Old Part
Retirement	2,160	52	16	0	0
New Full[a]	17	13	0	65	0
New Part[a]	13	2	0	0	78
Old Full	285	30	82	2,116	0
Old Part	41	25	1	0	317
Total 5,313					

[a] Job started is new in period before.

sample consists of either a single, full-time employment spell or a single retirement spell. Approximately 20% of the sample held two or more full-time jobs during the sample period. Part-time work is also fairly common with around 20% of the sample holding at least one part-time job.

Job choices are strongly related to age in the estimation sample as is shown in Table II. Only 20% of observations from 55 to 59 are retired while nearly 65% of observations over 65 are retired. The relative share of part-time work also increases strongly with age. Less than 5% of the observed jobs are part-time for individuals between 55 and 59, while over 40% of jobs are part-time for individuals 65 and over.

The patterns of transitions between job states are shown in Table III. As one would expect, the majority of observations are not changes in job states. Approximately 15% of jobs are not continued in the following year, and only 5% of retirement years are followed by employment spells. Despite these figures, there seems to be substantial movement between job states with nearly 800 cases of job switching and over 300 cases of working after a retirement spell in the full sample.[12] These cases of re-entry into jobs after retirement are examined more carefully in Tables IV and V. These tables show the proportion of retirement spells followed by jobs, broken down by age of respondent and

[12] Other retirement papers such as Diamond and Hausman (1984) and Gustman and Steinmeier (1986) do not allow for this possibility.

TABLE IV

JOB STATUS IN YEAR FOLLOWING RETIREMENT BY AGE OF INDIVIDUAL
ESTIMATION SAMPLE

		Proportion in Job State Following Year		
Age	Number of Observed Retirement Years	Retirement	Full-Time Work	Part-Time Work
55–59	159	.868	.113	.019
60–64	616	.937	.052	.011
65–69	920	.993	.002	.004
70 +	533	.996	0	.004

TABLE V

JOB STATUS IN YEAR FOLLOWING RETIREMENT BY LENGTH OF RETIREMENT SPELL
ESTIMATION SAMPLE

		Proportion in Job State Following Year		
Length of Retirement Year	Number of Observations	Retirement	Full-Time	Part-Time
1 year	384	.865	.115	.021
2 years	288	.969	.017	.014
3 years	256	.985	.008	.008
4 + years	1,400	.998	.001	.001

length of retirement spell respectively. The rates of re-entry into jobs decline significantly with both age and retirement tenure.

The significant decline of employment rates after retirement by age suggests that there may be a difference between an "unemployment" state and a "retirement" state. We take two measures to reduce any problems caused by omitting a separate "unemployment" state. First, as mentioned before, the estimation subsample is composed only of individuals who are 55 and older in 1966. Second, brief periods of unemployment that are surrounded by employment spells at the same firm are counted as "working" rather than "retired."

In conclusion, it appears that, while the majority of the sample follows a traditional employment pattern of a long term, full-time job followed by retirement, there is a substantial amount of job switching both to full-time and part-time jobs. Some re-entry into jobs after retirement spells is observed but may be caused partially by the grouping together of unemployment and retirement states.

4. ESTIMATION

4.1. *Model Specification*

Before we can proceed with estimation, we must specify the wage equations and the error distributions in more detail and show which parameters can be identified. The $\bar{w}$ equations are specified as

$$(4.1)\qquad \bar{w}_i(t,s,X)=\sum_{j=1}^{6}X_j(t)\bar{b}_{ij}+(t-s)\bar{b}_{i7}$$

for $i=F,P,R$ where the six exogenous characteristics $X_j(t)$, $j=1,2,\ldots,6$, are (1) a constant, (2) education, (3) race, (4) age, (5) age squared, and (6) health (and $\bar{b}_{R7}=0$). The $\tilde{w}$ equations are specified as

$$(4.2)\qquad \tilde{w}_i(t,s,X)=\sum_{j=1}^{6}X_j(t)\tilde{b}_{ij}+(t-s)\tilde{b}_{i7}$$

for $i=F,P,R$ where $\tilde{b}_{R7}=0$. As noted before, ε_i, $i=F,P,R$, has an extreme value distribution with variance $\pi^2\tau^2/6$. The other errors μ have a three factor structure:

$$(4.3)\qquad \begin{aligned} \mu_F(s)&=\bar{\mu}_F(s) &&=\sigma_1\eta_I+\sigma_2\eta_F+\sigma_3\eta_s,\\ \mu_P(s)&=\bar{\mu}_P(s)+\tilde{\mu}_P(s) &&=\sigma_1\eta_I+\sigma_2\eta_P+\sigma_3\eta_s,\\ \mu_R&=\bar{\mu}_R+\tilde{\mu}_R &&=\sigma_1\eta_I+\sigma_2\eta_R,\\ &&\tilde{\mu}_F(s)&=0, \end{aligned}$$

where η_I, η_F, η_P, η_R, and η_s, $s=1,2,\ldots,T^{**}$, are independent, identically distributed, standard normal random variables and σ_1, σ_2, and σ_3 are standard deviations. This is similar to the error structure in Miller (1984).

Equation (4.3) represents the structure of the errors known to the individual but unobserved by the econometrician. These represent random components of wages, nonwage job benefits, and tastes for leisure. The σ's are parameters to be estimated, and the η's are random effects that must be integrated out.

The first random effect η_I is a person-specific error, constant over the individual's lifetime. It has no effect on job status choices but affects wages. There is a status-specific random effect for each state: η_F, η_P, and η_R. These measure unobserved preferences an individual has for full-time work, part-time work, or retirement. The last set of random effects, η_s, $s=1,2,\ldots,T^{**}$, are time-specific and represent job-specific opportunities that arrive once a period.[13]

Even though the η's are independent, the μ's are not because η_I, η_F, η_P, and η_R are constant over an individual's life. In addition, an individual retains η_s as long as he holds the job that he began at time s. This factor structure for the μ's allows for unobserved heterogeneity. The model could also incorporate other error structures.[14] Flexible error structures pose no problem when using simulation estimators. The only limitation on the structure of the μ's (required

[13] The first job-specific error term, whether it is $\mu_F(1)$, $\mu_P(1)$, or μ_R, should not be distributed as a normal random variable because it represents the error connected to a job that was optimally chosen. We experimented with methods to control for this "initial conditions" effect. We found these methods unsatisfactory but also found that they had little effect on the other coefficients.

[14] See Miller (1984) and Pakes (1986) for other dynamic discrete choice models that allow flexible error structures.

for computational reasons) is that they be considered fixed, known constants in the individual's optimization problem, thus allowing use of equation (2.8) in computing value functions.

4.2. *Identification*

The data used to estimate the model are full-time wages and discrete, job state choices. The full-time wage is specified as

$$(4.4) \qquad w_F = \overline{w}_F + \overline{\mu}_F + \varepsilon_F .$$

This identifies $\bar{b}_{Fj}$, $j = 1, 2, \ldots, 7$, in equation (4.1).

The probability of choosing a particular job choice conditional on the μ's has a multinomial logit form because of the extreme value distribution of the ε's. For example, the probability of an individual, retired in year t, remaining retired in $t+1$ is

$$(4.5) \qquad \exp\{\overline{V}(R, t+1, s)/\tau\} / \big[\exp\{\overline{V}(R, t+1, s)/\tau\} + \exp\{\overline{V}(F, t+1, t+1)/\tau\} + \exp\{\overline{V}(P, t+1, t+1)/\tau\}\big]$$

where $\overline{V}$ is specified in equation (2.8). As is well known, discrete choice data will identify value functions only relative to some base. Thus, we treat the value of a new full-time job $V(F, t, t)$ as the base. Differencing equation (2.1) from $V(F, t, t)$ identifies $\tilde{b}_{F7}$ and the cost of moving c. Differencing equation (2.2) from $V(F, t, t)$ identifies $b_{Pj}^* = \bar{b}_{Pj} + \tilde{b}_{Pj} - \tilde{b}_{Fj}$, $j = 1, 2, \ldots, 6$, and $b_{P7}^* = \bar{b}_{P7} + \tilde{b}_{P7}$. Similarly, differencing equation (2.3) from $V(F, t, t)$ identifies $b_{Rj}^* = \bar{b}_{Rj} + \tilde{b}_{Rj} - \bar{b}_{Fj} - \tilde{b}_{Fj}$, $j = 1, 2, \ldots, 6$.

Variance terms, τ, σ_1, σ_2, and σ_3, are identified by second moments of full-time wages and discrete choices. The discount factor β is technically identified by the discrete choice data (conditional on τ) because it is the coefficient on the expected value of the optimal choice next period. However, we have chosen to estimate two models instead, one with fixed $\beta = 0$ and the other with fixed $\beta = .95$.[15]

4.3. *Method of Simulated Moments*

The estimation method used is the method of simulated moments (MSM) developed by Pakes and Pollard (1989) and McFadden (1989). The alternatives, maximum likelihood estimation or method of moments, are not feasible because transition probabilities involve integrating each conditional transition probabil-

[15] In preliminary experimentation, the estimation algorithm would not converge when we attempted to estimate β.

ity (conditional on all μ's for a specific person) over the joint density of the μ's. This requires integrating a $T^* + 4$ dimensional integral numerically.[16]

Let D_{nt} be a five-element, random vector of dependent variables for individual n at time t. The first four elements represent discrete choices, and the fifth is a full-time wage.[17] Let d_{nt} be the realization of D_{nt}; for the first four elements, $d_{nti} = 1$ if choice i was chosen, and the fifth element is the observed full-time wage.

The random variable D_{nt} is conditioned on relevant history. For the discrete choice elements of D_{nt}, relevant history is the job state individual n occupied at $t-1$ and when he accepted that position, $s(t-1)$. For the wage element of D_{nt}, relevant history is the job state occupied at t and when that job began, $s(t)$. Let the vector H_{nt} denote the vector of relevant history for D_{nt} element by element.

A method of moments estimator would require us to construct differences between d_{nt} and $E(D_{nt}|H_{nt})$.[18] However, $E(D_{nt}|H_{nt})$ is difficult to evaluate analytically because it depends upon a high dimensional integration over the joint distribution of μ and ε. Instead $E(D_{nt}|H_{nt})$ is simulated.

Let $D_{nt}^*(\theta, \mu) = E_\varepsilon[D_{nt}|\mu, H_{nt}]$ where θ is the vector of parameters to estimate. The discrete choice terms of D_{nt}^* are multinomial logit probabilities. The wage term of D_{nt}^* is

$$\bar{w}_F(X_{nt}, t, s(t)) + \mu_F + \rho E(\varepsilon_F|\mu, H_{nt}) \tag{4.6}$$

where ρ is a flexibility factor that is related to the correlation between true wage errors and job utility flows.[19] The last term of equation (4.6) is equal to

$$\rho\tau\left\{\gamma + \ln\left[\sum_j e^{\bar{V}_{Fj}/\tau}\right]\right\}, \tag{4.7}$$

where γ is Euler's constant and $\bar{V}_{Fj} = \bar{V}_j - \bar{V}_F$. Let $P_{nt}^*(\theta, \mu) = \Pr[H_{nt}|\mu]$. This is a relatively complicated sum of products of transition probabilities, each element of the sum corresponding to a different path from the initial condition to a condition at t consistent with H_{nt}. Let $\psi(\mu)$ be the joint density of μ. Then the $E(D_{nt}|H_{nt})$ is

$$\bar{D}_{nt}(\theta) = \frac{\int D_{nt}^*(\theta, \mu) P_{nt}^*(\theta, \mu) \psi(\mu)\, d\mu}{\int P_{nt}^*(\theta, \mu)\psi(\mu)\, d\mu}. \tag{4.8}$$

[16] See Lerman and Manski (1981) for a discussion of simulated maximum likelihood estimation and Avery, Hansen, and Hotz (1983) for a discussion of a method of moments estimator for a discrete choice model.

[17] For periods when a choice is not available or a full-time wage is not observed, the relevant element of D_{nt} is set to zero.

[18] The notation $E(D_{nt}|H_{nt})$ means that each element of D_{nt} is conditioned on its relevant history and not the relevant history of other elements of D_{nt}.

[19] If the model is correctly specified, ρ should equal 1 because the same error ε is specified in wage and discrete choice equations.

This can be simulated by

$$(4.9)\qquad \tilde{D}_{nt}(\theta)=\frac{\frac{1}{R}\sum_{r=1}^{R} D_{nt}^{*}(\theta,\mu_r)P_{nt}^{*}(\theta,\mu_r)}{\frac{1}{R}\sum_{r=1}^{R} P_{nt}^{*}(\theta,\mu_r)},$$

where μ_r, $r=1,2,\ldots,R$, are R iid pseudo-random variables with density ψ.[20] Simulated residuals are constructed as $\tilde{e}_{nt}(\theta)=d_{nt}-\tilde{D}_{nt}(\theta)$. The method requires an unbiased simulator of $\overline{D}_{nt}(\theta)$. In fact, $\tilde{D}_{nt}(\theta)$ is not an unbiased estimator of $\overline{D}_{nt}(\theta)$ because the denominator needs to be simulated as well as the numerator. For this application, however, the bias is small.[21]

The MSM estimator of θ_1 is

$$(4.10)\qquad \tilde{\theta}_1=\underset{\theta_1}{\arg\min}\,\tilde{e}(\theta)'ZZ'\tilde{e}(\theta)/N$$

where N is the sample size, $\tilde{e}(\theta)$ is the vector of $\tilde{e}_{nt}(\theta)$ stacked over periods and then individuals, Z is a matrix of instruments, and θ_1 is a subset of θ. The instruments must satisfy: (1) $E[\tilde{e}(\theta)|Z]=0$ and (2) $Z'\tilde{e}_{\theta}/N$ converges to a nonsingular matrix where $\tilde{e}_{\theta}=\partial\tilde{e}/\partial\theta$. Instruments are chosen to be approximations to

$$(4.11)\qquad \frac{\int[\partial\ln D_{nt}^{*}(\theta,\mu)/\partial\theta]P_{nt}^{*}(\theta,\mu)\psi(\mu)\,d\mu}{\int P_{nt}^{*}(\theta,\mu)\psi(\mu)\,d\mu}$$

for the elements of Z corresponding to discrete choices[22] and

$$(4.12)\qquad \frac{\int[\partial D_{nt}^{*}(\theta,\mu)/\partial\theta]P_{nt}^{*}(\theta,\mu)\psi(\mu)\,d\mu}{\int P_{nt}^{*}(\theta,\mu)\psi(\mu)\,d\mu}$$

for elements of Z corresponding to full-time wages. The procedure used is to simulate equations (4.11) and (4.12) and then form instruments by regressing them on exogenous regressors z. The exogenous regressors z are needed because equations (4.11) and (4.12) depend upon job tenure which is endoge-

[20] The μ's are drawn once at the beginning and then held fixed throughout the estimation process.

[21] We performed some small Monte Carlo experiments to measure the bias. In the first experiment we could not reject a null hypothesis that the bias is zero. In the second, based on a simplified special case of $\tilde{D}_{nt}(\theta)$, we determined that the bias is only large when there is both much variation in P_{nt}^{*} and D_{nt}^{*} across draws of μ_r and the correlation between P_{nt}^{*} and D_{nt}^{*} is close to -1. For applications where the bias is large, an unbiased estimator of $\overline{D}_{nt}(\theta)$ could be simulated using acceptance-rejection methods.

[22] These are approximations to the optimal instruments for method of moments. See McFadden (1989, p. 1004). They also satisfy conditions (iii) and (iv) of Theorem (3.3) of Pakes and Pollard (1989).

nous. The exogenous regressors are constructed by the following procedure:

(1) Conditioning only on initial job state and an initial value of θ, denoted θ^0, use the dynamic programming problem to simulate job state sequences and, for each sequence that results in a chosen job state, calculate its probability along with required derivative terms.

(2) Calculate approximate job state probabilities by summing over sequences resulting in that job state, and calculate approximate derivatives by taking probability weighted sums of derivative terms.

(3) Use the approximate values calculated in (2) in equations (4.11) and (4.12) to construct z.

The preceding procedure is used to estimate

$$\theta_1 = \{\bar{b}_{Fj},\ j=1,2,\ldots,7;\ b^*_{Pj},\ j=1,2,\ldots,7;\ b^*_{Rj},\ j=1,2,\ldots,6;\ c;\ \rho\}$$

conditional on $\theta_2 = \{\sigma_1, \sigma_2, \sigma_3, \tau\}$.

The covariance parameters θ_2 are estimated conditional on θ_1 using MSM on simulated second moments. Let $\hat{C} = \Sigma_n \tilde{e}_n \tilde{e}_n'/N$.[23] The expected value of $\hat{C}$ is

$$(4.13) \quad \bar{C} = E_\mu\{E_\varepsilon[D_{nt} - \bar{D}_{nt}][D_{nt} - \bar{D}_{nt}]'|\mu\} + E_\mu\{[D^*_{nt} - \bar{D}_{nt}][D^*_{nt} - \bar{D}_{nt}]'|\mu\}.$$

The first term represents the covariance due to variation in extreme value errors. The diagonal wage terms inside the first brackets { } are always equal to $\pi^2\tau^2/6$ which does not depend upon μ. The off-diagonal wage terms are all zero because the extreme value errors are independent. The discrete choice terms have the usual multinomial logit covariance structure, and the covariance of wages with discrete choices are zero (because wage observations are conditional on full-time work). The second term represents the covariance due to variation in the job-specific normal errors. While it cannot be evaluated analytically, it can be simulated by measuring variation in predicted wages and simulated discrete choice probabilities over draws of μ. The simulated covariance matrix $\tilde{C}$ provides an unbiased estimator of $\bar{C}$.

Let $\tilde{c}$ be the vector of nonrepeating elements of $\tilde{C}$ and $\hat{c}$ be the vector of nonrepeating elements of $\hat{C}$. Define second moment residuals as $\tilde{g}(\theta) = \hat{c}(\theta) - \tilde{c}(\theta)$. The MSM estimator of θ_2 conditional on θ_1 is

$$(4.14) \quad \tilde{\theta}_2 = \underset{\theta_2}{\arg\min}\, \tilde{g}(\theta)'\tilde{g}(\theta).$$

The algorithm used to estimate the whole model is to: (1) Initialize θ at θ^0. (2) Construct instruments conditional on θ^0. (3) Estimate $\tilde{\theta}_1$ conditional on θ_2^0 from equation (4.10). (4) Estimate $\tilde{\theta}_2$ conditional on $\tilde{\theta}_1$ from equation (4.14). (5) Set

[23] The actual construction adjusts for missing observations and does not condition on relevant history. It is infeasible to condition on relevant history in constructing $\hat{C}$ because there are not enough observations with the same relevant history. Thus the construction of $\tilde{e}_{nt}$ considers all possible paths from the initial state to the state occupied at t. The same point applies to the terms in equation (4.13).

θ^0 to $(\tilde{\theta}_1, \tilde{\theta}_2)$ and return to step (2) unless $\tilde{\theta}$ has converged.

4.4. *Asymptotic Properties*

In this section we derive the asymptotic properties of our estimator $\tilde{\theta}$ using the results of Pakes and Pollard (1989). Denote $M(\theta)$ as the moment condition that must equal zero at the true value of θ, denoted θ^*. For θ_1, $M(\theta) = Z'e(\theta)$, and for θ_2, $M(\theta) = g_\theta(\theta)'g(\theta)$, where $e(\theta)$ and $g(\theta)$ are true (not simulated) residuals. Let $M_N(\theta)$ be the simulation of $M(\theta)$ as described in Section 4.3,[24] and let $\tilde{\theta}_N$ minimize $\|M_N(\theta)\|$ where $\|\cdot\|$ is the Euclidean norm.

We assume the identification conditions:

$$\sup_{\|\theta - \theta^*\| > \alpha} \|M_N(\theta)\|^{-1} = O_p(1) \tag{4.15}$$

and that $M(\theta)$ is differentiable at θ^* with a derivative matrix M_θ of full rank. Our simulators $M_N(\theta)$ satisfy the Lipschitz condition of Lemma (2.13) of Pakes and Pollard. The identification conditions and the Lipschitz condition are sufficient for Theorems 3.1 and 3.3 of Pakes and Pollard to apply. Theorem 3.1 states that $\tilde{\theta}$ converges in probability to θ^*, and Theorem 3.3 states that

$$\sqrt{N}(\tilde{\theta} - \theta^*) \sim N\left[0, (M_\theta^* M_\theta)^{-1} M_\theta' \Sigma M_\theta (M_\theta' M_\theta)^{-1}\right]. \tag{4.16}$$

The asymptotic covariance matrix can be written as $Q^{-1}\Omega Q^{-1}$ where

$$Q = \operatorname{plim}\left[Z'\tilde{e}_\theta / N\right] \tag{4.17}$$

and

$$\Omega = \operatorname{plim}\left[Z'(I \otimes \tilde{C}) Z / N\right]. \tag{4.18}$$

The efficiency of the MSM estimator relative to the method of moments estimator is $R/(R+1)$ where R is the number of pseudo random draws used in simulation (McFadden (1989)).

5. RESULTS

We estimated a dynamic model ($\beta = .95$) and a static model ($\beta = 0$) using ten independent random draws of μ. The coefficient estimates and their standard errors are reported in Table VI. The retirement and part-time coefficients, b_{Rj}^* and b_{Pj}^*, $j = 1, 2, \ldots, 7$, from Section 4.2, represent the effect of each variable on the relative value of being retired or part-time employed, respectively. For example, using the dynamic model, one more year of education decreases the value of retirement (relative to a new full-time job) by \$76.20 per year and increases the relative value of a part-time job by \$42.70 per year. The full-time coefficients, $\bar{b}_{Fj}$, $j = 1, 2, \ldots, 7$, from Section 4.2, represent the effect of each variable on full-time wages. For example, one more year of education increases annual, full-time wages by \$358.

[24] In particular, for θ_1, $M_N(\theta) = Z'\tilde{e}(\theta)$ and, for θ_2, $M_N(\theta) = \tilde{g}_\theta(\theta)'\tilde{g}(\theta)$.

TABLE VI
MODEL ESTIMATES

Choice	Variable	$\beta = 0$	$\beta = .95$
Retirement	Constant	$-120(25.8)^*$	$-30.8(5.26)^*$
	Education	$-3.82(1.34)^*$	$-.762(.254)^*$
	Race	14.5(10.8)	1.56(2.20)
	Age	$24.6(3.61)^*$	$4.29(1.03)^*$
	Age**2	$-.630(.238)^*$	.0389(.0292)
	Health	$30.0(3.77)^*$	$11.1(1.00)^*$
Part-Time	Constant	$-118(33.6)^*$	$-32.7(13.6)^*$
	Education	4.15(2.20)	.427(.387)
	Race	$-8.58(22.5)$	$-1.51(4.55)$
	Age	$16.8(7.70)^*$	2.38(2.71)
	Age**2	$-.500(.459)$	$-.0365(.134)$
	Health	4.16(8.81)	$5.93(2.46)^*$
	Tenure	7.50(12.5)	.613(.671)
Full-Time	Constant	$-103(12.3)^*$	$-47.5(7.47)^*$
	Education	$3.37(1.08)^*$	$3.58(.798)^*$
	Race	$-21.5(9.15)^*$	$-7.07(6.68)$
	Age	$-7.08(1.90)^*$	$2.22(1.23)^*$
	Age**2	$.0242(.127)^*$	$-.223(.0842)^*$
	Health	$-10.4(3.03)^*$	$-5.77(2.04)^*$
	Tenure (b_{F7})	$1.37(.329)^*$	.398(2.25)
	Tenure (b_{F8})	1.04(.690)	.0872(.0618)
Other variables	Cost of changing	$190(24.9)^*$	$120(6.48)^*$
	τ	35.0	26.4
	σ_1	.118	46.7
	σ_2	54.2	.0739
	σ_3	138.0	6.72
	ρ	$-.376(.276)$	.123(.170)

Notes: (1) Numbers in parentheses are standard errors. (2) Starred entries have an asymptotic t statistic greater than 2. (3) Wages are reported in annual wages $\div 100$. (4) Retirement and part-time coefficients should be interpreted as changes in the value of that state relative to the value of full-time employment. (5) Age is measured as true age minus 55. (6) Estimation results use 10 independent random draws of μ.

For both models, bad health decreases the relative value of a full-time job, education decreases the relative value of retiring, and age increases the relative value of retiring. Furthermore, both models explain job attachment mainly through the high cost of changing jobs ($12,000 for the dynamic model and $19,000 for the static model). Finally, neither model has an estimated flexibility factor ρ close to one. This indicates that the relationship between unobservables in value functions and in wage equations is probably misspecified.

There are some significant differences between the two models. First, the dynamic retirement and part-time coefficients are generally smaller than the static coefficients. This occurs partly because the dynamic τ is smaller than the static τ, but mostly because the effect of a variable in the dynamic model on relative values depends upon the future as well as the present. For example, one more year of education increases the value of working not only through its contemporaneous effect but also through its effect on the relative increase in future flows associated with work.

TABLE VII

SUMS OF SQUARED RESIDUALS

Weights	Dynamic	Static	Naive
Unweighted	1,197	1,461	2,971
Ω_{95}^{-1}	79,494	83,376	
Ω_0^{-1}	12,425,901	11,419,581	
Ω_0^{-1} (Outliers excluded)	597,847	608,507	

Notes: (1) There were 500 individuals with 14545 degrees of freedom. (2) The naive model includes only estimated constants and variance terms. (3) Ω_{95} is the dynamic covariance matrix, and Ω_0 is the static covariance matrix. (4) There were eight outliers when Ω_0 was used, each with a squared residual greater than 100,000, and two with squared residuals greater than 1,000,000.

Second, the correlation structure of the errors is different in the two models. The dynamic model errors are smaller, and the two largest factors, σ_1 (person-specific effect) and τ (extreme value effect), do not increase job attachment. On the other hand, the static model has a large job-specific effect σ_3 and a large status-specific effect σ_2; both of these increase job attachment. The reason for the difference is that the static model needs large effects that remain over the tenure of a job to counterbalance temporary surprises. These are not needed in the dynamic model because the individual places less value on temporary events.

In order to compare our estimates to most other estimates in the literature, we need to compute derivatives of log wages. Using the average full-time wage ($6,904), we find the effect of education, age, and bad health to be 5.2%, -2.9% (for a 64 year old man), and -16.7% respectively. These results are similar to the results in Gordon and Blinder (1980) and Gustman and Steinmeirer (1986).[25]

There is no straightforward way to test the two models against each other because they are not nested models.[26] Thus we have developed a set of relevant weighted sums of squared residuals reported in Table VII. Though we cannot compute the distribution of any of the statistics, they suggest that the dynamic model fits the data better than the static model.

The first row of Table VII displays unweighted sums of squared residuals for three models. The dynamic and static models are compared with a naive model in which only constant terms and variance parameters are estimated. The dynamic model performs best, and both dynamic and static models fit much better than the naive model.

The first row is of limited value because the squared residuals are neither identically nor independently distributed. The last six statistics attempt to

[25] Gordon and Blinder's effects are 3%, -2.6% and -10% respectively. Gustman and Steinmeier's effects are 7.8% for education in excess of 12 years, 4.1% for education less than 12 years, and 0% for experience (for a male with 44 years of experience). Gustman and Steinmeier do not control for bad health.

[26] If we could estimate β, then there would be no issue.

correct this by weighting with an inverse covariance matrix of the residuals.[27] Unfortunately, the covariance matrix depends upon the particular set of model coefficients used to estimate covariances. Statistics computed using the covariances from the dynamic model result in a weighted sum of squares from the static model which is 1.04 times as large as the dynamic weighted sum of squares.[28]

When the covariance matrix used is constructed from the static coefficients, the static model performs better than the dynamic model. However, this is because of eight observations with very large squared errors. These eight outliers, all with values of 100,000 or more, account for about 95% of the total weighted sum of squares. Without these eight outliers, the dynamic model performs slightly better than the static model. Thus, these informal tests suggest that the dynamic model fits the data better.

The choice of model has a large effect on simulated behavior. Consider the hazard rates in Figure 1. The three curves represent simulated hazard rates (using 50 draws) of a 55 year old, healthy, white man with ten years of education and ten years of experience. The top and bottom curves use the static and dynamic model respectively and are quite different.

The static and dynamic models differ by having different estimated coefficients and also by having different discount factors. In order to separate these effects, the hazard rate was simulated using the static model coefficients but with $\beta = .95$; this is shown as the middle curve in Figure 1. As shown in this figure, less than half of the discrepancy in hazard rates is due to different discount rates; the rest is caused by different coefficients. The increased volatility of the middle curve is caused by the combination of the large fixed effects in the static model with the increased weight on the future from $\beta = .95$. This combination causes responses to be more sensitive to high draws of the simulated errors. The "jumpiness" of the curve is caused by the variability in the number of extreme draws over time. In order to smooth out this curve, substantially more than 50 independent draws would be needed.

More subtle differences in behavior between two models can be seen in Figure 2. The curves with solid circles are simulated hazard rates for the same individual in Figure 1 except that he becomes permanently disabled at age 64. The curves with open circles are simulated hazard rates for the same individual except that he becomes only temporarily disabled at age 64. For the static model, hazards are exactly the same until age 65 because there is no difference in the individual's environment until age 65. On the other hand, the dynamic hazards show that the individual both anticipates future disability and discounts

[27] The covariance matrix used is block diagonal, each block corresponding to a particular year for a particular individual. Discrete choice correlations across time were set to zero because each residual was conditional on past relevant history. Correlations of wages across time were set to zero because we had no routine to invert an 85×85 covariance matrix. Each individual's covariance matrix is simulated with 50 independent draws.

[28] We do not know what the distribution is of the ratio of the sums of squares. Thus, we can make no statements about the significance of the ratios in Table VII.

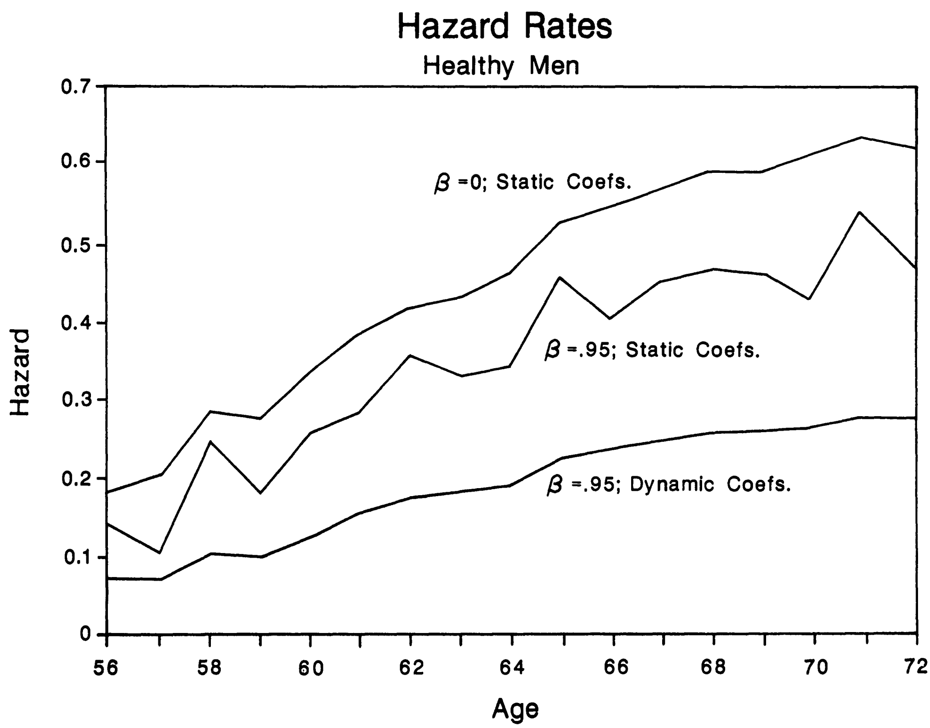

FIGURE 1

the effect of temporary disability. Furthermore the effect of permanent disability on behavior is much greater in the dynamic model than in the static model. This is solely because an individual in the dynamic model values future employment less relative to retirement than does the individual in the static model (note that the dynamic retirement health estimate is smaller than the static retirement health estimate).

The health change experiments are only suggestive in nature due to the extreme assumptions about knowledge of future health status. Nevertheless, we interpret the results in these two figures as evidence that it is important to explicitly include the dynamics in the estimation procedure of a dynamic model. In reality, neither model is correct because individuals know some information about their future health and recognize that it will affect the value of future choices (i.e., the static model is incorrect), but they do not have perfect information about their future health (i.e., the dynamic model is incorrect). In fact, reality should be somewhere between the two extremes of the models estimated in this paper.

The last issue concerns the importance of the unobserved heterogeneity caused by individual and job specific effects. The curves in the first two figures are hazard rates for an average 55 year old man with ten years of education and

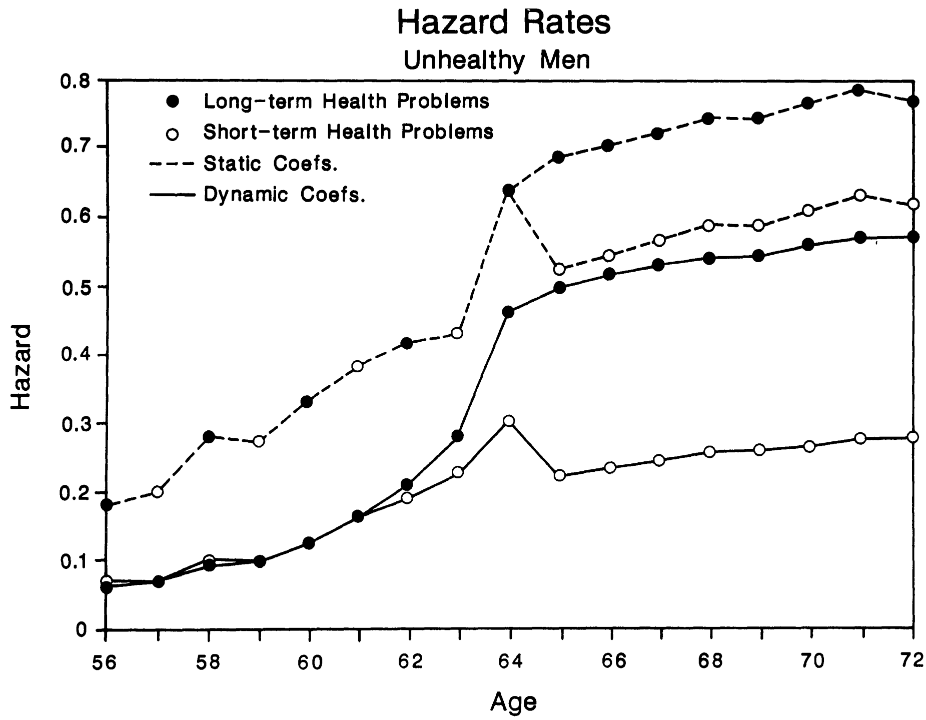

FIGURE 2

ten years of experience. But the hazard rates at, for example, age 64 are not those in the figures because the sample of individuals still working at age 64 enjoyed better than average job opportunities than the sample of 55 years old still working.[29] Figure 3 shows the effects of unobserved heterogeneity on estimated hazard rates.[30] The solid curves are different simulations of the healthy individual of Figure 1 and the permanently unhealthy individual in Figure 2. The dashed curves are hazard rates for the same two experiments except that the draws used to simulate the curves are weighted by the probability that they would have resulted in the individual still being employed at each point in the curve. For example, at age 64, the unweighted healthy curve is

$$\frac{1}{R}\sum_{r=1}^{R} D_{ntR}^{*}(\theta, \mu_r), \tag{5.1}$$

[29] See Flinn and Heckman (1982) for a discussion of the effects of unobserved heterogeneity on estimated duration dependence.

[30] These use the dynamic coefficients.

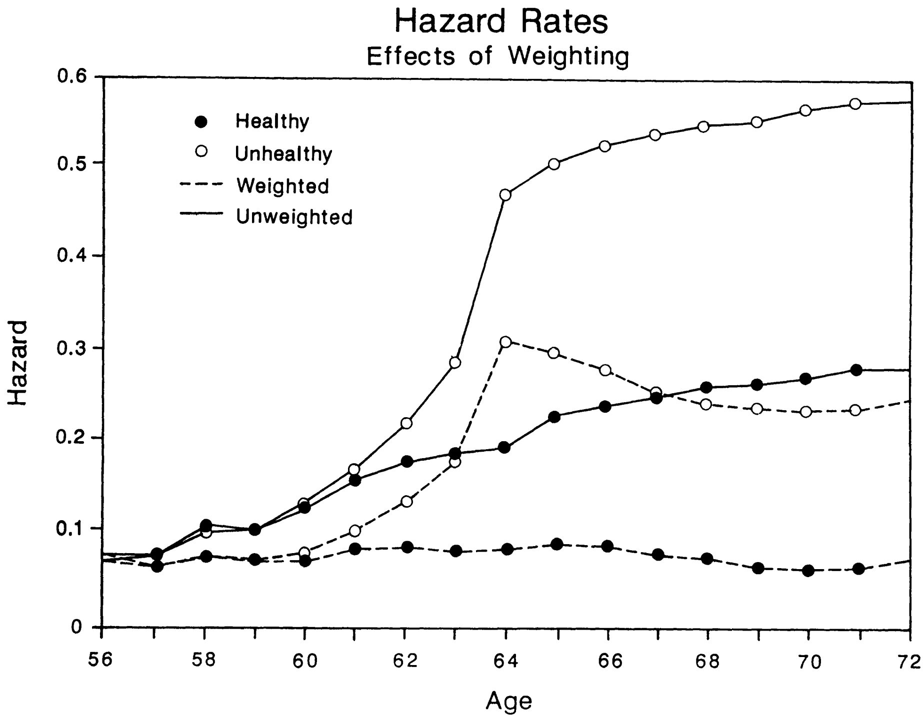

FIGURE 3

and the weighted healthy curve is

$$(5.2) \qquad \frac{\frac{1}{R}\sum_{r=1}^{R} D_{ntR}^{*}(\theta,\mu_r)P_{nt}^{*}(\theta,\mu_r)}{\frac{1}{R}\sum_{r=1}^{R} P_{nt}^{*}(\theta,\mu_r)}$$

where D_{ntR}^{*} is the retirement probability defined in D_{nt}^{*} of equation (4.9). Equation (5.2) is what one would observe if he or she took a random sample of 55 year olds and observed them over 17 years. It includes the negative duration dependence caused by unobserved heterogeneity.

The magnitude of the bias in Figure 3 shows that the inclusion of the serially correlated errors has important effects on interpreting hazard rate data. This paper has found some evidence of significant unobserved heterogeneity. Though it is significantly more difficult to estimate a dynamic model with unobserved heterogeneity,[31] if it is an important component of an individual's behavior, it needs to be modelled.

[31] Note its absence in innovative work such as Rust (1984, 1987, 1988) or Eckstein and Wolpin (1986).

6. CONCLUSION

Our results indicate that dynamics play an important part in labor force participation decisions of mature men. Our dynamic model performs better than our static model using crude statistics based on sums of squared residuals. It is important to explicitly include the dynamics of the model in the estimation procedure. Using estimates from a procedure based on a static model to predict behavior in a dynamic model can lead to poor predictions.

Whether or not dynamic models fit the data, they capture a component of reality that static models cannot address. Most economic models of retirement have dynamic components in them.[32] Interpreting dynamic models using coefficients estimated from static models can lead to severe interpretation errors as seen in Figure 1. Furthermore, one can test how important dynamics are only by allowing for their possibility in an econometric model.

It is also important to allow for unobserved person and job-specific effects of job values in a labor force participation model whether the model is dynamic or static. Most of the previous literature on estimation of dynamic models has not allowed for these effects because standard estimators (maximum likelihood or method of moments) become infeasible to compute.[33] The method of simulated moments allows for much more flexible correlation structures with little loss in efficiency.

The model estimated in this paper is flawed in that it does not include the Social Security system and it has unreasonable assumptions about the information available to each individual. The first flaw explains the lack of a high hazard rate into retirement at ages 62 and 65. The second explains the results in Figure 2. Both of these flaws can be corrected using the same modelling and estimation procedure suggested in this paper, though at an increased cost. We feel that this methodology has great promise in estimating structural, dynamic models.

Federal Reserve Board, Washington, D.C., U.S.A.

and

Department of Economics, University of Virginia, Charlottesville, VA 22901, U.S.A.

Manuscript received February, 1988; final revision received January, 1990.

REFERENCES

AVERY, ROBERT B., LARS PETER HANSEN, AND V. JOSEPH HOTZ (1983): "Multiperiod Probit Models and Orthogonality Condition Estimation," *International Economic Review*, 24, 21–35.

DIAMOND, PETER, AND JERRY HAUSMAN (1984): "The Retirement and Unemployment Behavior of Older Men," in *Retirement and Economic Behavior*, ed. by Henry Aaron and Gary Burtless. Washington, DC: Brookings Institute.

[32] See, for example, Gustman and Steinmeier (1986).

[33] Only Miller (1984) has successfully used maximum likelihood estimation in a dynamic discrete choice model with unobserved heterogeneity errors. Pakes (1986) uses the simulation estimator of Lerman and Manski (1981) which is computationally expensive, especially in a problem with a large number of choices to be made per period such as this one.

DUBIN, JEFFREY, AND DANIEL MCFADDEN (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

ECKSTEIN, ZVI, AND KENNETH WOLPIN (1986): "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience," paper prepared under a contract with the Employment and Training Administration, U.S. Dept. of Labor, under the authority of the Comprehensive Employment and Training Act.

——— (1989): "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models," *Journal of Human Resources*, 24, 562–598.

FLINN, CHRISTOPHER, AND JAMES HECKMAN (1982): "Models for Analysis of Labor Force Dynamics," in *Advances in Econometrics*, Vol. 1, ed. by G. Rhodes and R. Basemann. New London, CT: JAI Press.

GORDON, ROGER, AND ALAN BLINDER (1980): "Market Wages, Reservation Wages, and Retirement Decisions," *Journal of Public Economics*, 14, 277–308.

GUSTMAN, A., AND T. STEINMEIER (1986): "A Structural Retirement Model," *Econometrica*, 54, 555–584.

HECKMAN, JAMES, AND THOMAS MACURDY (1980): "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies*, 47, 47–74.

LERMAN, STEVEN R., AND CHARLES F. MANSKI (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Charles F. Manski and Daniel McFadden. Cambridge, Mass.: MIT Press.

MACURDY, THOMAS (1981): "An Empirical Model of Labor Supply in a Life Cycle Setting," *Journal of Political Economy*, 89, 1059–1085.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

MILLER, ROBERT (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.

PAKES, ARIEL (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.

PAKES, ARIEL, AND DAVID POLLARD (1989): "The Asymptotics of Simulation Estimators," *Econometrica*, 57, 1027–1058.

RUST, JOHN (1984): "Maximum Likelihood Estimation of Controlled Discrete Choice Processes," SSRI Working Paper 8407, University of Wisconsin, 1984.

——— (1987a): "A Dynamic Programming Model of Retirement Behavior," National Bureau of Economic Research Working Paper 2470.

——— (1987b): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1034.

STERN, STEVEN (1989): "Measuring the Effect of Disability on Labor Force Participation," *Journal of Human Resources*, 24, 361–395.

UNITED STATES, DEPARTMENT OF HEALTH AND HUMAN SERVICES, SOCIAL SECURITY ADMINISTRATION, OFFICE OF THE ACTUARY (1982): *Life Tables for the United States: 1900–2050*, SSA Publication No. 11-11534. Washington, DC: Government Printing Office.

WOLPIN, KENNETH (1984): "An Estimated Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 852–874.