NBER WORKING PAPER SERIES

AN ANALYSIS OF THE HEALTH AND RETIREMENT STATUS OF THE ELDERLY

Robin C. Sickles

Paul Taubman

Working Paper No. 1459

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1984

This research was supported by a National Institutes of Health grant to the National Bureau of Economic Research. We wish to thank C. A. Knox Lovell, Janet Meininger and Janice Shack-Marquez for useful comments. We would also like to thank Alan Mathios for his valuable research assistance. The research reported here is part of the NBER's research program in Health Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #1459
September 1984

An Analysis of the Health and Retirement Status of the Elderly

ABSTRACT

In this paper we specify and estimate a structural limited dependent variable model with which we study both the health and retirement status of the elderly. Standard linear estimators, which assume that these variables are continuous, are not appropriate and categorical estimation techniques are preferred. Our model differs from previous work in that we have longitudinal data and random effects that are correlated over time for different individuals. The problem is made more complicated because there is sample truncation, which could potentially bias coefficient estimates, since approximately twenty percent of the individuals in our sample die. We outline the full information maximum likelihood estimator for such a model and implement it in our empirical analysis. With our structural estimates we analyze, among other things, the degree to which endogeneously determined health status affects the probability of retirement and how changes in social security benefits and eligibility for transfer payments modify both healthiness and the demand for leisure.

Robin C. Sickles
Paul Taubman
Department of Economics
University of Pennsylvania
3718 Locust Walk/CR
Philadelphia, PA 19104

(215) 898-7708

Substantial empirical interest has been focused recently on the question of how health and retirement decisions are related. The empirical work to date typically has used a reduced form model and, most recently, nonlinear specifications such as the multinomial logit (Anderson and Burkhauser (1983)) or hazard rate model (Hausman and Wise (1983a)) have been used. Standard linear estimators, which assume that these variables are continuous, are not appropriate and categorical estimation techniques are preferred. The aforementioned studies have generally found that health and retirement are not independent and that health and Social Security wealth have significant effects on the retirement decision.

In this paper we focus on the efficient estimation of a structural model of the health and the retirement decisions of the elderly. Our model differs in part from previous work in that we have longitudinal data and random effects that are correlated over time for different individuals. Efficient estimation of a structural system of limited dependent variables such as ours in the context of a panel data set has to our knowledge not been implemented. The problem is made more complicated because there is a twenty percent death rate and hence sample truncation which could potentially bias coefficient estimates.

We build the full information maximum likelihood estimator from the univariate results of Butler and Mofitt (1982) and use it in our empirical analysis. With our structural estimates, we can analyze, among other things, the degree to which the endogeneously determined health status affects the probability of retirement and how changes in Social Security benefits and eligibility for transfer payments modify both healthiness and the demand for leisure. The studies cited above addressed these issues in part but either used inefficient estimators or based their findings on reduced form models. Another study which focused primarily on retirement (Hanoch and Honig (1983)) did not control for random effects. By controlling for heterogeneity we can be more precise and can examine the extent to which ignoring heterogeneity could bias parameter estimates. Furthermore, we can test whether or not a systems estimator has any empirical appeal by a direct test of the significance of the covariance parameter.

The plan of the paper is as follows. Section 1 discusses the economic model of joint health and retirement status and the variables that enter into it. Section 2 outlines the statistical model. Section 3 discusses the data set used in the analysis and reviews our estimation results. Section 4 concludes.

1. The Model

Our model contains both a health and a retirement equation. The retirement equation is based on the assumption that an individual maximizes a utility function given by:

$$(1.1) \qquad U = U(C,L,H)$$

where C is consumption, L is leisure, and H is health.[1]
Health is included to account for pain and suffering and shifts in tastes, e.g., some activities may be less desirable if you have a physically limiting health problem such as arthritis.

Equation (1.1) is maximized subject to a budget constraint and a health production function given in equations (1.2) and (1.3).

$$(1.2) \qquad Y_t = w_t(T-L) + r_tA_t + X_t = P_cC_t + P_H Z_t$$

$$(1.3) \qquad H_t = F(Age_t,\ Z_t,\ O_t))$$

where w is the hourly wage rate, T is total time (hours) in time period t, r is the return on financial investments, A is the amount of financial assets, X

[1]Although our model is a long-run static one, some elements of other periods are allowed to enter via discounting future benefits.

is other sources of income including earnings of a spouse, P_c is the price of consumption, P_H is the price of medical care, Z is the amount of medical care, and 0 is job and personal characteristics.

In principle it is possible to find the first order conditions for a maximum and to obtain demand equations for health, for hours of work and leisure, and for consumption. However, since the theoretical model of retirement is well known and since the information available to us does not allow us to obtain labor supply elasticities, we will merely specify the arguments entering each of the equations.[2]

The standard retirement model asserts that an individual compares the utility generated from working versus that from fully or partially retiring. Utility differs in these situations because working is unpleasant and because income differs. If retired, the individual may be eligible for Social Security benefits and pensions. If he works, he may forfeit all or a part of these benefits and pensions but he receives a wage or salary. Future Social Security and pension payments may also be changed if retirement is postponed. Thus we wish to include in the retirement equation the benefits if retired permanently, the expected change in the benefits if retirement is postponed a year, wage earnings if the person works, and other sources of income.

While anecdotal stories exist that suggest retirement per se causes bad health via boredom, we are not aware of any firm evidence that retirement debases health. To the extent that economic decisions are made in a rational fashion, retirement should not directly modify the unobservable health stock. Moreover, Ekerdt (1983) in a detailed study based on medical examinations found retirees' health deteriorated no more than that of a

[2]See Parsons (1980), Boskin (1977) or Quinn (1977).

control group of nonretirees. Thus we specify a triangular model with health affecting retirement but with no feedback from retirement to health.[3]

The general specification of the health equation is based on the work of Anderson and Burkhauser (1983), Grossman (1972), Lee (1982), Taubman and Rosen (1982), and Taubman and Sickles (1984). The unobservable health stock is endogeneously determined and can be augmented by investment in health services or depreciated by the environment of the work place. The health stock differs across individuals and families and is determined in part by: social and demographic factors such as education, longest occupation, race and age; the degree to which an individual is able to gain access to information on available health services which we proxy by marital status, number of children, and education; and ability to pay for health services for which we include income, assets, spouse's income, and pension and social security benefits.

We omit from the analysis in year t those already dead. However we include these same people in earlier periods if alive. During the 8 years the survey spans, more than 20% of the initial respondents died and there is a selectivity problem whose solution is discussed below.

2. Statistical Model

Our statistical model is an extension of the single equation limited dependent variable model of Heckman (1981). The longitudinal nature of the data set is accommodated by using a conventional error components specification (Balestra and Nerlove (1966)) in which heterogeneity between individuals is modeled as a random effect. There are two equations in our system -- one which links health status to the retirement decision and one

[3]A statistical test of this maintained hypothesis -- which is accepted -- is discussed in Section 3.

which models changes in the unobservable health stock. Since we argued earlier that these are jointly determined, a systems estimator would be expected to yield more efficient estimates than limited information techniques. In this section we outline both full information (FIML) and limited information maximum likelihood (LIML) estimators.

The system can be written as

$$(2.1) \qquad y_{it}^{*(1)} = x_{it}^{(1)} \beta_1 + \varepsilon_{it}^{(1)}$$

$$y_{it}^{*(2)} = \gamma_2 y_{it}^{(1)} + x_{it}^{(2)} \beta_2 + \varepsilon_{it}^{(2)} \qquad \begin{array}{l} i = 1, \ldots, N \\ t = 1, \ldots, T \end{array}$$

where

$$(2.2) \qquad \varepsilon_{it}^{(j)} = \mu_i^{(j)} + v_{it}^{(j)} \qquad (j = 1, 2)$$

and where

$$E[\varepsilon_{it}^{(j)} \varepsilon_{ks}^{(1)}] = \begin{array}{l} \sigma_\mu(j) + \sigma_v(j) \text{ for } j=1,\ i=k,\ t=s \\ \sigma_\mu(j) \text{ for } j=1,\ i=k,\ t\neq s \\ \sigma_{v_1 v_2} \text{ for } j\neq 1,\ i=k,\ t=s \\ 0 \text{ elsewhere .} \end{array}$$

Here $x_{it}^{(1)}$ and $x_{it}^{(2)}$ are $(1xk_1)$ and $(1xk_2)$ vectors of exgoneous variables, β_1 and β_2 are conformable vectors of structural coefficients, $y_{it}^{*(1)}$ and $y_{it}^{*(2)}$ are scalar unobserved dependent variables whose observed counterparts are $y_{it}^{(1)}$ and $y_{it}^{(2)}$; γ_2 is a scalar coefficient for the right-hand-side endogeneous observable, and $\varepsilon_{it}^{(1)}$ and $\varepsilon_{it}^{(2)}$ are the errors in the two equations which are decomposed by the rule in (2.2). The unobserved are linked to the observed by the following rules:

$$(2.4) \qquad y_{it}^{(1)} = j \text{ if } A_{j-1}^{(1)} - x_1\beta_1 < \varepsilon_{it}^{(1)} < A_j^{(1)} - x_1\beta_1,\ j=1,\ldots,4$$

with $A_0^{(1)}$, $A_4^{(1)}$ normalized at $-\infty$, $+\infty$ respectively, and

$$(2.5) \qquad y_{it}^{(2)} = j \text{ if } A_{j-1}^{(2)} - x_2\beta_2 < \varepsilon_{it}^{(2)} \leq A_j^{(2)} - x_2\beta_2, \ j=1,2$$

with $A_0^{(2)}$ and $A_2^{(2)}$ normalized at $-\infty$, $+\infty$ respectively.

Equation (2.4) is the polytomous probit with ordered responses and (2.5) is a binary probit. We use the standard normalization that $\sigma_v(j) = 1$, $j=1,2$. We furthermore assume that the errors are distributed normally, with joint density $f(\varepsilon_{it}(1), \varepsilon_{is}(2);\theta)$ where $\theta = (\beta_1', \beta_2', \gamma_2, \sigma_\mu^{(1)}, \sigma_\mu^{(2)}, \sigma_{v_1 v_2})$. We assume that the parameter space is compact, that the likelihood $L_N(y_{it}^{*(1)}, y_{is}^{*(2)}; \theta)$ based on N panels is a continuous function of θ for every set $y^* = \{y_{it}^{*(1)}, y_{js}^{*(2)}\}$, that $(1/N)\log L_N(y^*; \theta)$ converges to a function $Q(\theta)$ almost surely uniform for every θ, and that this function has a unique maximum at the true parameter point. We furthermore assume that the log likelihood function is three times differentiable, and that the absolute value of the third derivative is bounded by some function with finite expectation (Cramer (1946)). Based on these assumptions the maximum likelihood estimate of θ is consistent and has the limiting distribution given by $\sqrt{N}\,(\hat{\theta}-\theta) \rightarrow N(0,\ell^{-1})$ where ℓ is the Fisher information matrix. A computational issue arises when implementing FIML since calculation of the joint probabilities of observing differing configurations of health-retirement states for the same individual over the T time periods is problematic if the number of time periods is large. Our data set contains five biennial periods. Since the calculation of the joint retirement-health states for an individual at time t requires two dimensional integration, the calculation of the set of retirement-health states for an individual over (at most) five dependent time periods requires ten dimensional integration.

Numerical methods for handling such problems are available (Clark (1964)) but are both computationally burdensome and have an approximation error which is difficult to bound.

Fortunately, the evaluation of multi-dimensional integrals is made much simplier by the particular form for the correlation pattern of disturbances implied by the variance components model. This point has recently been made by Butler and Moffit (1982) for simpler univariate probit models. For the joint model a similar approach can be taken.

We first define the domains of integration over which the various functionals are evaluated. Since there are eight possible configurations -- $\binom{4}{1}\binom{2}{1}$ -- there are eight domains $(D_1,\ldots,D_8)$. Let R be the domain corresponding to a particular configuration of health-retirement states and examine the joint probability of observing this generic configuration for the ith individual at time t

$$\int_R\int dF = \int_R\int f(\varepsilon_{it}^{(1)},\ \varepsilon_{it}^{(2)})\ d\varepsilon_{it}^{(1)}\ d\varepsilon_{it}^{(2)} \quad .$$

For the ith individual the joint probability of observing the T health-retirement states is

$$\int_R\ldots\int dF = \int_R\ldots\int f(\varepsilon_{i1}^{(1)},\ldots,\varepsilon_{iT}^{(1)},\varepsilon_{i1}^{(2)},\ldots,\varepsilon_{iT}^{(2)})d\varepsilon_{i1}^{(1)}\ldots d\varepsilon_{iT}^{(2)}$$

$$= \int_R\ldots\int \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(v_{i1}^{(1)},\ldots v_{iT}^{(1)},v_{i1}^{(2)},\ldots,v_{iT}^{(2)}/\mu_i^{(1)},\mu_i^{(2)}\ )\ \cdot$$

$$g(\mu_i^{(1)},\mu_i^{(2)})\ d\mu_i^{(1)}d\mu_i^{(2)}dv_{i1}^{(1)}\ \ldots dv_{iT}^{(2)}$$

$$= \int_{-\infty}^{\infty} h_1(\mu_i^{(1)}/\mu_i^{(2)})[\ \int_{-\infty}^{\infty} h_2(\mu_i^{(2)})[\prod_{t=1}^{T}\ \int_R\int f(v_{it}^{(1)},v_{it}^{(2)}/\mu_i^{(1)},\mu_i^{(2)})\ \cdot$$

$$dv_{it}^{(1)} dv_{it}^{(2)}] d\mu_i^{(2)}] d\mu_i^{(1)}$$

$$= \int_{-\infty}^{\infty} h_1(\mu_i^{(1)}/\mu_i^{(2)})[\int_{-\infty}^{\infty} e^{-z^2} h_3(z)dz]\ d\mu_i^{(1)}$$

$$= \int_{-\infty}^{\infty} e^{-w^2} h_4(w)dw \ . \tag{2.6}$$

This last expression can be evaluated using the Hermite Integration formula

$$\int_{-\infty}^{\infty} e^{-w^2} h_4(w)dw = \sum_{j=1}^{G} w_j h_4(w_j)$$

where G is the number of evaluation points and $h_4(w_j)$ is h_4 evaluated at w_j. The evaluation is done in a nested fashion with evaluation of the bivariate distribution function, for which highly accurate algorithms exist, left as the only numerical burden.

Denote this last expression as P_i. Then the likelihood function is

$$L_N(\theta;y^*) = \sum_{i=1}^{N} P_i(\theta;y^*) \ . \tag{2.7}$$

We turn our attention now to the way in which sample truncation is handled and to the calculation of estimates of θ . The former point must be addressed since sample attrition occurs over time as people die and, for a particular individual, the conditional probability of an observed health-retirement state (with health state possibly deceased), given that the person was alive in the previous period, is a function of the joint probability of not having died in all previous time periods. Sample truncation can be handled by conditioning the joint probability of retirement-health states for an individual on the joint probability that the individual was alive in the previous periods. If the individual dies, he is removed from the sample in the next period. Call this joint probability $P_i^*(\theta;y^*)$ and denote the modified likelihood function as $L_N^*(\theta;y^*)$. Under the conditions outlined above the FIML estimates have the limiting distribution

$$\sqrt{N}\ (\hat{\theta}-\theta) \rightarrow N(0,\lim[\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \log L_i^*}{\partial \theta'} \frac{\partial \log L_i^*}{\partial \theta}]).$$

To generate LIML estimates, first estimate the health equation using the single equation analogue of our bivariate model. Next estimate the quasi-reduced form for the retirement equation conditional on the observed health state using maximum likelihood. Finally concentrate the likelihood function (2.7) with respect to all parameters except $\sigma_{v_1 v_2}$ and maximize it with respect to the single covariance parameter. Note that all these calculations utilize $P_i^*(\theta;y^*)$ instead of $P_i(\theta;y^*)$. The LIML estimates can also be used in a single Newton-Raphson iteration to yield consistent and asymptotically efficient estimate of θ.

3. Data, Variables, and Estimation Results

The data come from the Retirement History Survey (RHS) which contains five biennial panels taken during the period 1969 through 1977 and individually matched records of Social Security earnings beginning in 1951. The sample contains about 8500 men who were heads of households in 1969.[4] It contains objective health information such as data on death and hospitalization and subjective information such as how your health compares with others of the same age and how it has changed over time.

Our dependent variable indicator for retirement is constructed in the

[4]Because of potential differences in the way women and men perceive their health, we deleted 2500 women heads of household from our analysis and put off estimating a model with a fully interactive female status dummy.

following fashion. One question asked in the RHS is "are you presently working part time, full time, or are you retired?" We chose to use the working full-time/not-working-full-time dichotomy instead of either deleting those who are semi-retired or introducing a new category for retirement status. We have examined models in which we use semi-retired as another category and found comparable results. We rely on either the retired or not retired results for simplicity and comparability with most other studies.

Turning to the health measure, we find a number of empirical studies which conclude that health affects retirement decisions. (See footnote 2.) Many of these studies have used one of two health measures. The first is the answer to a question like "Does your health limit your ability to work or to get about?" For people who retired prior to their 65th birthday, this type of question allows, and perhaps invites, the subject to cite health limitations as a socially acceptable reason. A second health measure frequently used is whether or not a person died within some follow-up time interval. Used alone this is also a far from ideal measure. Deaths from accidents and from diseases which strike swiftly and which would not be preceded by pain, suffering, and loss of ability in earlier years are quite different than those caused by chronic diseases. Long-term debilitating illnesses would not be accurately modeled with the early post-sample death measure while future accidents and the like probably don't affect retirement calculations.

Our study uses a subjective measure of health status which we think is superior to the aforementioned measures. The RHS solicites answers to the question "How does your health compare with that of others of the same age?" The possible responses are better, same, or worse. The public health literature suggests that subjective ratings by the elderly are highly correlated with the arguably more objective physician ratings (Ferraro

(1980)), Mossey and Shapiro (1982)). Also, in the RHS, the people who report themselves in worse health are twice as likely to die in a four year span as those in better health (Taubman and Rosen (1982)), and generally display the properties one would expect in a health production function (Taubman and Rosen (1982), Asher (1984)). Taubman and Sickles (1984) have used the objective/subjective health variable to analyze the health effects of the Supplemental Security Income (SSI) program with quite reasonable results.

To these subjective rankings we add a fourth category -- deceased. While the RHS records a person's death when they learn of it during an attempted reinterview, they often don't obtain this information. We rely instead on files from Social Security who record this information to stop paying benefits to the deceased, begin paying survivors their benefits, and to justify paying burial allowances. Recent work by Duleep (1983) indicates that in recent years these files are extremely accurate in obtaining death information. The Social Security data go through 1979. Both year and month of death are given. We have cross-checked the RHS files against Social Security's information up to the 1977 survey date. We found 2 instances where the RHS lists the person as dead but the other file doesn't. We also found that if Social Security lists an individual as dying between two surveys, the RHS either lists that individual as having no response or the RHS indicates that the respondent is the surviving spouse.

Before turning to the results we should point out that a number of different specifications of the joint health-retirement model were considered. We first examined a model in which the endogenous variables both entered in their unobservable forms as right-hand-side variables. Although we considered only consistent two-step estimators of these models (Mallar (1977)) our results were not at all supportive of this type of fully latent structure. Furthermore, the effect of the unobservable work effort variable on the health stock was negligible and insignificant. The question of whether or not a mixed model, in which observed/unobserved endogenous variables appear on the right hand side, was also considered. Because of the categorical nature of both endogenous variables, mixed structures or structures in which just the observed counterparts of both endogenous variables appear on the right hand side would be forced by coherency conditions to be triangular (Heckman (1978), Gourieroux, Laffont, and Moufort (1980)). Therefore, only a model in which health affects the mean of retirement or one in which retirement effects the mean of health (specified as unobservables or as observable counterparts) are empirically relevant. These are inherently nonnested models. Two-step procedures were again used to examine competing structures. The effect of observed retirement status on the health stock was (as with its latent counterpart) negligible and insignificant. For this reason and the reasons cited in section 1, we focus on the triangular system in which health determines the mean level of retirement propensity but in which no direct feedback is permitted from retirement to health. Unexplained effects can certainly cause unexplained variations in the two endogenous variables to be correlated and this provides FIML with an efficiency gain over LIML.

Due to computational constraints, a random sample of 808 people was selected from the roughly 8500 people in our original sample. The means and standard deviations, calculated from the panel, of the variables used in the analysis are presented in Table 1.

Social Security benefits are the benefits one would expect to receive if retirement begins in the respective year. It is computed using covered earnings taken from each person's Social Security record, which is part of the

RHS, and then replicating Social Security's rules. Thus we first calculated each person's Average Monthly Earnings (AME). This was accomplished by using the respondent's earnings since 1951, which were truncated at the maximum allowable earnings level. The five lowest years of income are dropped and the sum of the remaining incomes is divided by the number of months worked. The resulting AME is then used to compute the Primary Insurance Amount (PIA) based on the tables in the Social Security Handbook. These account for inflation and therefore change over the 1969-1977 sample period. Once the PIA was computed, the benefits total was determined on the basis of PIA and marital status. By using benefits available rather than those paid to actual retirees, we avoid an obvious selection problem. It should be pointed out that since benefits are increased by 50 percent if the individual is married, the effect of marital status on both retirement and health will depend in part on the benefits' coefficient.

Income from assets is the sum of yearly income generated from the value of assets: stocks, bonds, life insurance annuities, etc. Pensions are not incorporated in this variable but are included separately. The gain from postponing retirement is calculated by taking earnings in 1969 (In each respective year these earnings are inflated by the CPI.) plus the gain in Social Security benefits from postponing retirement one additional year -- discounted to the averaged expected lifetime of the individual based on age -- less the Social Security benefits the individual would have received. As Mitchell and Fields (1983) have noted, this variable may be positively correlated with retirement since the substitution effect (away from leisure) may dominate the income effect (toward leisure). Average income from assets was lower than average Social Security benefits but close to average pension income while spouse's earnings were greater than all three.

We included a dummy variable to indicate whether or not the person was eligible in 1975 or 1977 for Supplement Security Income (SSI), which began in 1974, and interacted this variable with a time trend to identify changes in the health stock over time for SSI eligibles. In an earlier study with a somewhat different model of health, Taubman and Sickles (1984) found that those who were eligible to receive SSI in 1975 or 1977 were in worse health in 1969 than those who would not meet the eligibility criterion, but the differential narrowed over time and became insignificant.

As shown in Table 1 most of the men are married although widowers make up about 5% of the person-year observations. In our subsample all the men happened to have been married at some point in their life and the omitted category is thus divorced/separated. The most common longest occupation was as a skilled worker with the omitted category of unskilled workers accounting for roughly 25% of the sample.

An interesting problem arises because eligibility for Social Security's old age benefits only occurs at age 62. A 60 year old could calculate the value of his future Social Security benefit stream and obtain an unsecured loan against it or run down existing assets, but, since (nonhousing) assets are small, this may be difficult if capital markets are imperferct. We allow for these difficulties by including in the retirement equation a pre age 62 dummy variable.[5] We now turn to the estimation results.

The retirement equation, presented in Table 2, is familiar to economists although it has a few novel variables as well as some interesting quantitative results. The advantages of not retiring in a particular year are given by the "gain from postponing retirement" variable. It is the most significant of the

---

[5]This is not perfectly colinear with a set of time dummies since in 1969 and 1971 some people are not 62.

income variables and its coefficient an order of magnitude larger than the asset, spouse's earnings, or pension income coefficients. We can easily translate the raw coefficients into marginal probabilities at the sample means by scaling the estimated coefficient by the normal density evaluated at the estimated mean of the index describing $y^{*(2)}$.[6] If the gain from postponing retirement is increased one standard deviation from its sample mean ($739), then the probability of retiring is reduced by 0.044. Similar increases in income from assets, spouse's earnings, and pension income increase the probability of retirement by about .02, .03, and .035 respectively.

Of the occupation variables, only the self-employed dummy seems to have any significant explanatory power and the effect is quite sizeable: self-employment reduces the probability of retiring by 0.18.

During the 1970's retirement benefits paid by Social Security increased substantially faster than inflation. To some extent benefits grew because of secular growth in wage rates which help determine an individual's primary insurance amount and benefits. However, to a large extent, the benefits increased because of two legislated changes. One was the institution of Supplemental Security Income (SSI) to the elderly on welfare.[7] SSI gave money to those on welfare and made them eligible for Medicaid. The other change was the provision of overgenerous protection against changes in the CPI. The indexing provisions were technically deficient because both the benefit schedules and earnings histories, to which the benefit schedules were applied, were shifted with the CPI. Our structural model allows us to examine the consequences of these changes. In preliminary analysis SSI was found to have

---

[6]The mean index for the retirement equation is 12.9 and the normal density corresponding to $A_1^{(2)} - \bar{X}^{(2)}\hat{\beta}^{(2)}$ is 0.391.

[7]To some extent SSI replaced state based Old Age Assistance but on average it increased benefits substantially.

an insignificant and second order effect in the retirement equation which is why it is excluded in our final results. However, Social Security benefits (embedded in our gain from postponing retirement variable) were quite significant and had a relatively large effect. The first major change in benefits occurred from January 1971 to September 1972 and was about 13 percent in real terms (Leimer and Lesnoy (1983)). The second major change -- basing benefits on wage-indexed earnings -- occurred in September 1977. According to figures compiled by Summers (1982), the ratios of primary benefits for an "average-earnings" man retiring at age 65 to earnings in the year before retirement were 34.3, 39.4, 40.7, 43.6, 45.5 for the years 1971, 73, 75, 77, 79. The effect of these Social Security reforms on the probability of retirement seems to be quite small, amounting to only .84, .34, 1.90, and 1.25 percentage point increases in the respective years. Although our results do not directly tell us what effect these reforms had on the age of retirement, they indicate that its effect is rather limited and are consistent with the retirement age effects found by Hausman and Wise (1983b) and Fields and Mitchell (1984).

Turning to the age variable, we find that it is, not surprisely, quite significant and highly correlated with retirement. Based on its coefficient, an individual of 62 is almost 17% less likely to retire than an individual of 64. The dummy for being less than 62 has a highly significant negative coefficient. Finding the effect of age less than 62 is complicated because changing the variable results in an obvious change in age. If we look at the effect of aging one year from 61 to 62, then the probability of retiring increases by almost .31 while aging one year from 60 to 61 increases the probablity by almost .08. It would seem that there is evidence for either substantial imperfection in capital markets or a fairly high discount rate for the people in our sample.

Years of schooling have an important bearing on the retirement decision. The better educated retire later in their lives. For example, a college educated male would, at age 64, be almost .12 less likely to retire than a high school graduate. While human capital models often assume that the more educated have the same length of career as the less educated to make the analytics more tractable, there is no necessary reason for this to occur. However, it may well reflect the differential work activities of the more educated which are less affected by aging.

Married and widowed males are both less likely to retire than those who are divorced or separated although the coefficient for the widowed category is not very significant. The actual effect of marital status on retirement is confounded by the gain from postponing retirement variable. Since benefits one would receive if one retired are increased by 50% if the individual is married, there is an obvious interaction between the two variables. At average levels of the gain from postponing retirement, married males are about 16% less likely to retire than the divorced or separated, with a t-statistic of -3.37. The widowed are less likely to retire than the divorced or separated by about .08.

We next focus attention on the health variable which is significant at the 99% level. Previous work has not considered the form in which health affects the retirement decision. That is, should health status enter the retirement equation in its unobservable form or should an observable counterpart be used? This issue of appropriate specification is in principle a testable hypothesis. Because of the triangular nature of the system, both variables could be included in the retirement equation and conventional t-tests could be carried out. However, as a practical matter the inclusion of both health measures will result in severe multicollinearity and render the test rather powerless. A somewhat different strategy which we use does not require a composite model to test the hypothesis. Furthermore, it has the attractive property that the general specification of the model, as opposed to just the particular hypothesis concerning the appropriate health measure, can be tested. Under the null hypothesis that the measured health status variable is appropriate, FIML on the system will result in consistent and asymptotically efficient estimates. Under the alternative that the retirement equation is misspecified, the FIML estimates of the health equation will be inconsistent because of specification error in the correlated retirement equation. However, in this case the LIML estimates of the health equation will still be consistent. Thus the Hausman-Wu test can be utilized. The χ^2 statistic for this test is 15.6 while $\chi^2_{21,0.05} \doteq 31$ suggesting that our specification is appropriate. Further support for our specification and the main reason for the relatively low test-statistics, is that the estimated correlation between equations is only -.101 with a t-statistic of -1.79.

Because the health variable is endogenous, we cannot manipulate the other explanatory variables since health status is changed by the same variables we are holding constant in the retirement equation. If we view the change in health status in an *ex ante* sense then a movement from a poor to a good health status reduces the probability of retirement and thus increases expected average earnings by almost .21. Add to this figure the potential reduction in medical expenditures owing to better health, and there is a substantial real income gain from lower morbidity.

Before moving to the health equation estimates we note that the random effects are sizeable, accounting for almost 1/3 the total variation in the total error, and highly significant. Heterogeneity is quite evident in the retirement decision.

We now turn to the health equation, results for which are presented in Table 3. Recall that the variable is scaled so that higher numbers indicate worsening health. We present our results as we did with the retirement equation and in general compare states of better health with health same as others of your age.[8]

Focusing first on the economic variables, we see that the only first-order effects come from social security benefits and from pension income. For example, an annual increase in these variables by $10,000 would increase the probability of being in better health by about .16 and .09 respectively.[9] Eligibility for transfer payments from the Supplemental Security Income program are not highly significant although the point estimates provide evidence of the same sort found by Taubman and Sickles (1984).[10] Those who were eligible for SSI in 1975/77 were in worse health in 1969 than those who would not meet the eligibility criteria but the differential narrowed over time. From 1969-1977 the probability of being in better health for SSI eligibles increased by about .16 and the probability of dying <u>fell</u> almost .65.[11]

---

[8] Other binary comparisons are easily made by appropriately modifying the thresholds since the only quantatitive differences are the $A^{(1)}$'s (and the average value across states for the explanatory variable whose effect we are analyzing). The mean index for the health equation is 2.064 and the normal density associated with the probability of being in better health is 0.349. This will be the scale factor in analyzing the marginal probabilities associated with the raw coefficients.

[9] It is possible that long term ill health has reduced labor market activity and earnings which determine the benefits. Since we do allow for individual specific effects in our equations, we don't think this is the cause of the correlation.

[10] At average levels of benefits the t-statistic associated with the joint hypotheses that SSI has no effect is 1.34.

[11] One qualification on the SSI results should be noted. Eligibility for SSI in 1975 or 1977 is not completely known for those who died in 1974 or earlier. For those who died prior to 1975 we do have information on whether they were receiving state assistance when the survey starts and they were

The other statistically significant coefficients are on the age, number of dependents, education and longest occupation variables. Remembering that health is scaled such that poorer health receives a higher number, the age effect is not surprising even if people compare themselves to others of the same age. This means that more older people are dead and/or that people compare themselves to the median rather than the mean person of the same age.

It is generally argued that the more educated are brighter, are better equipped to make decisions, make more informed decisions, and adapt new products more quickly. Thus it is not surprising that the more educated are in better health, ceteris paribus. An increase in education completed from 12 to 16 years would raise the probability of being in better health by about .05.

The omitted longest occupation in our sample is unskilled labor who are in worse health than people in the other occupations. We can not determine if this occurs because their job worsens their health, because less healthy people are more likely to work as unskilled laborers, or because poorer people invest less in health preserving regimes. The probability of being in better health is .16 lower for unskilled laborers than, for example, those who had been in management positions.

An increase from two to three dependents increases the probability of being in better health by almost 2%. There are several possible reasons for this outcome. First a number of people have argued that larger social networks lead to better health with people exchanging information on health and doctor quality.[12] Second, healthy (unhealthy) parents may beget healthy

alive. These people were eligible for SSI. However, it seems that we are still understating the number of eligibles who died prior to 1975. Thus the health of the SSI group between 1969 and 1975 should be worse than our numbers would indicate, meaning that the estimated SSI dummy and the relative improvement over time for the SSI group are probably understated.

(unhealthy) children. The unhealthy children may die early. Moreover unhealthy parents may chose to have fewer children because of their low income and energy levels.

We can also examine the direct effect of legislated changes in Social Security benefits on the healthiness of the aged and, through the health and gain from postponing retirement variables, on their propensity to retire. Based on Summers' figures the total change from 1971-79 in the probability of being in better health due to the reforms was only .0199. The feedback from the health equation would lessen the probability of retiring by only about .2%, reducing the 2.8% direct increase in the probability of retiring due to the Social Security reforms of 1971-79 changes to about 2.6%.

Heterogeneity is significant and important in the health equation. Although the relative size of the random effects is smaller than with the retirement equation, random effects still contribute almost 30% to the total unexplained variation in the health stock.

A final comment should be made about the use of FIML over LIML estimation. The former is approximately an order of magnitude more cpu intensive than the latter. Controlling for heterogeneity in both the retirment and the health equations using direct controls and individual specific random effects seems quite adequate in reducing the correlation in unexplained variations to a small (-.101) and marginally insignificant (t-statistic = -1.79) level. It is not as clear that the substantial computational investment needed to carry out FIML on models of this sort is justified.

[12] See Asher (1982) for a survey on social networks' impact on health and some important evidence.

4. Conclusions

This study has focused on the structural estimation of a joint health-retirement model in which both sample truncation and error dependencies substantially complicate the implementation of an efficient estimator. We have been able to isolate the effect of perceived health status' on the retirement decision in a structural setting. Furthermore, we have performed several important policy simulations to see how the double indexing and increased transfer payments affected the retirement decision both directly and by way of modifications in the health status of the eligible individuals. Our results indicate that retirement decisions are strongly affected by health status, variables that change the shape and position of the income/leisure opportunity set, marital status, self employment status and education. We also find that those not yet eligible for Social Security status are far less likely to retire. This suggests that if part of the solution to the known future financing difficulties of the Social Security System involves raising the normal retirement age to 67 or 68, then a major policy decision is whether to leave 62 as the early retirement age or to raise it to 64 or 65. The latter change would induce more people to work longer and pay more taxes.

Our health equation results indicate that Social Security and pension payments have positive effects on healthiness. The other significant variables are number of dependents and longest occupation being unskilled.

We calculate that the planned and unplanned increases in Social Security benefits in the 1970's raised the probability of retirement by about .026 and increased the percentage in better health by almost .02.

We also find that random effects are quite important in both equations. However, due to the rather small and marginally significant estimated correlation between equations after we control for heterogeneity, efficiency gains from a FIML do not appear to be worth the substantial computational investment necessary for its implementation.

Table 1

Sample Summary Statistics

Variable	Mean	Standard Deviation
Age	64.3	3.31
Age Less Than 62	.215	.411
Black	.0592	.236
Married	.902	.297
Widowed	.0511	.220
Number of Dependents	.229	.713
Receiving SSI	.0275	.164
Years of Education	10.02	3.15
Longest Occupation		
Professional	.218	.409
Clerk	.0893	.285
Skilled Labor	.429	.495
Management	.150	.357
Self-Employed	.104	.305
Social Security Benefits	1716.	1642.
Income From Assets	1113.	3782.
Spouse's Earnings	2039.	4469.
Pension Income	1237.	3541.
Gain From Postponing Retirement	594.	739.
Health	1.93	.775
Retirement	1.55	.498

Table 2

Estimation Results
Retirement Equation

Variable	Coefficient	t-statistic
Age	.211	43.7
Age <62	-.585	-7.78
Black	.0501	0.31
Married	-.259	-2.05
Widowed	-.213	-1.27
Number of Dependents	-.0128	-0.37
Education	-.0771	-6.52
Professional	.0171	0.12
Clerk	.0289	0.18
Skilled Laborer	.0613	0.51
Management	.00286	0.02
Self-Employed	-.468	-6.61
Income From Assets	$.138 \times 10^{-4}$	2.17
Spouse's Earnings	$.169 \times 10^{-4}$	2.44
Pension Income	$.249 \times 10^{-4}$	2.14
Gain From Postponing Retirement	$-.151 \times 10^{-3}$	-4.63
Health	.263	3.13
Threshold 1	12.7	38.6
$\sigma_{\mu}^{(2)}$	.522	13.7

Table 3

Estimation Results
Health Equation

Variable	Coefficient	t-statistic
Age	.0448	7.18
Black	-.114	-1.00
Married	-.00209	-0.02
Widowed	-.145	-1.14
Number of Dependents	.0555	2.13
Education	-.0382	-4.27
Professional	-.337	-4.03
Clerk	-.349	-3.54
Skilled Laborer	-.296	-4.44
Management	-.467	-5.24
Self-Employed	-.0563	-0.85
SSI	4.00	1.18
SSI x Time	-.0540	-1.21
Income From Assets	$-.594 \times 10^{-5}$	-0.99
S.S. Benefits	$-.462 \times 10^{-4}$	-3.29
Pension Income	$-.266 \times 10^{-4}$	-4.09
Spouse's Earnings	$-.131 \times 10^{-5}$	-0.27
Threshold 1	1.55	3.66
Threshold 2	3.08	7.53
Threshold 3	4.08	9.76
$\sigma_\mu^{(1)}$	.375	19.6

References

Anderson, K., and R. Burkhauser: "The Effect of Actual Mortality Experience within a Retirement Decision Model," unpublished manuscript, 1983.

Anderson, K., R. Burkhauser, and J. Quinn: "Do Retirement Dreams Come True? A Rational Expectations Approach to the Retirement Decision," unpublished manuscript, 1983.

Asher, C.: "The Impact of Social Support Networks on Adult Health," Medical Care, 22 (1982), 349-359.

Balestra, P., and M. Nerlove: "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, 34 (1966), 585-612.

Bartel, A., and P. Taubman: "The Impact of Specific Diseases on Earnings," Review of Economics and Statistics, 61 (1979), 1-9.

Boskin, M.: "Social Security and Retirement Decisions," Economic Inquiry, 15 (1977), 1-15.

Butler, J. S., and R. Moffitt: "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," Econometrica, 50 (1982), 761-764.

Clark, C.: "The Greatest of a Finite Set of Random Variables," Operations Research, 9 (1964), 145-162.

Cramer, H.: Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.

Duleep, H.: "The Socioeconomic Determinants of Mortality: The Role of Income," paper presented at the American Economic Association Winter Meetings, December, 1983.

Ekerdt, D., et al.: "The Effects of Retirement on Physical Health," American

Journal of Public Health, 73 (1983), 779-783.

Ferraro, K.: "Self-Ratings of Health Among the Old and the Old-Old," Journal of Health and Social Behavior," 21 (1980), 377-383.

Fields, G., and O. Mitchell, "The Effects of Social Security Reforms on Retirement Ages and Retirement Incomes," NBER Working Paper No. 1348, May 1984.

Gourieroux, G., J. J. Laffont, and A. Monfort: "Coherency Conditions in Simultaneous Linear Equation Models with Endogeneous Switching Regimes," Econometrica, 48 (1980), 675-696.

Grossman, M.: "On the Concept of Health Capital and the Demand for Health," Journal of Political Economy, 80 (1972), 223-255.

Hanoch, G. and M. Honig: "Retirement, Wages, and the Labor Supply of the Elderly," Journal of Labor Economics, 1 (1983), 131-151.

Hausman, J., and D. Wise: "Social Security, Health Status, and Retirement," unpublished manuscript, NBER, March, 1983a.

________________________, "Retirement and Subsequent Work," Paper presented at the National Bureau of Economic Research Conference on Pensions, Labor, and Individual Choice, Dorado Beach, Puerto Rico, March 1983b.

Heckman, J.: "Dummy Endogeneous Variables in a Simultaneous Equation System," Econometrica, 46 (1978), 931-960.

___________: "Statistical Models for Discrete Panel Data," in Structural Analysis of Discrete Data with Econometric Applications, edited by C. Manski and D. McFadden. Cambridge: MIT Press, 1981.

Lee, L. F.: "Health and Wage: A Simultaneous Equations Model with Multiple Discrete Indicators," International Economic Review, 23 (1982), 199-222.

Leimer, D., and S. Lesnoy: "Social Security and Private Saving: An Examination of Feldstein's New Evidence," ORS Working Paper Series #31, U.S. Department of Health and Human Services, October, 1983.

Mallar, C.D.: "The Estimation of Simultaneous Probability Models," Econometrica 45 (1977), 1717-1722.

Mitchell, O., and G. Fields: "The Economics of Retirement Behavior," NBER Working Paper No. 1128, 1983.

Mossey, J., and E. Shapiro: "Self-Rated Health: A Predictor of Mortality Among the Elderly," American Journal of Public Health, 72 (1982), 800-808.

Parson, D.: "The Decline in Male Labor Force Participation," Journal of Political Economy, 88 (1980), 117-134.

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

Summers, L.: "Observations on the Indexation of Old Age Pensions," NBER Working Paper No. 1023, November, 1982.

Taubman, P., and R. Sickles: "Supplemental Security Income and the Health of the Poor", NBER Working Paper No. 1062, June, 1984.

Taubman, P., and S. Rosen: "Healthiness, Education and Marital Status," in Economic Aspects of Health, edited by V. Fuchs. Chicago: University of Chicago Press, 1982.