Econometric Analysis of Health Status Dynamics Among the British Elderly

Multivariate Markov Chain Models With Unobserved Heterogeneity[#]

Christian Salas

Health Services Management Centre, University of Birmingham[§]

Abstract

This paper provides confirmatory evidence that risk factors such as smoking, loneliness and low socioeconomic status remain significant at old age even after *simultaneously* controlling for past health, self-reporting errors and both observed and unobserved heterogeneity. No previous economic or epidemiologic study has controlled for all these features together. The health capital model is augmented with a Partial Adjustment Model to conceptualise the stochastic process governing health changes in older age-groups as a first-order Markov chain. This is estimated using self-assessed health data on the elderly in the British Household Panel Survey. New methods are used to tackle three issues: (1). the self-assessed health data give rise to 'ordered' states; this is exploited to yield parametric parsimony; (2). self-reporting errors are allowed for; (3). time-persistent unobserved heterogeneity invalidates the *first-order* model, and makes past health endogenous. Two ways of dealing with this are explored in a 'random effects' set-up: integration by Hermite quadrature, and a new Monte Carlo simulation approach.

Key words: markov chains, unobserved heterogeneity, health, elderly, risk factors
JEL Classification: C13, C33, C51, I12

[#] I am grateful to my colleagues Lisa Gold and David Parry for helpful comments on an earlier draft of this paper, although the usual disclaimer applies with regard to any remaining errors or omissions.
[§] Address for correspondence: Dr. Cristian P. Herrera-Salas, Lecturer in Health Economics, University of Birmingham, Health Services Management Centre, Park House, 40 Edgbaston Park Road, Birmingham B15 2RT. Tel.: (0121) 4143197. Fax: (0121) 4147051. Email: herrercp@hsmc.bham.ac.uk.

1. Introduction

This paper presents a dynamic multivariate analysis of health status transitions among elderly respondents in the 1991-1995 waves of the British Household Panel Survey. It provides confirmatory evidence that risk factors such as smoking, loneliness and low socioeconomic status remain statistically significant at old age even after *simultaneously* controlling for past health, self-reporting errors and both observed and unobserved heterogeneity. Previous dynamic analyses in the health economics literature have focused on nursing home residents using restrictive 'homogeneous' Markov chains which ignore heterogeneity across individuals (see Norton, 1992, and references therein). Portrait, Lindeboom and Deeg (1999) present multivariate analyses of health changes among institutionalised and non-institutionalised respondents from a Longitudinal Aging Study in Amsterdam (LASA), but their models are not dynamic in the sense of incorporating lagged health terms. Their results do not therefore show the relevance of variables like smoking, loneliness and socioeconomic status *after* controlling for past health.

Epidemiologic studies of the elderly have identified risk factors such as isolation, physical inactivity, extremes of body mass, poor nutrition, alcohol consumption and tobacco use, some using dynamic multivariate models.[1] However, none of these are based on formal behavioural models of health changes at old age, which are used in this paper to bring to light the effects of problems like *unobserved heterogeneity* and *endogeneity* of past health. These have not been discussed in any epidemiologic studies of the elderly.

The analysis in this paper is based on a formal behavioural model which suggests that health follows a stationary first-order Markov chain at old age, provided that there is no unobserved heterogeneity in transition rates. The model indicates that unobserved heterogeneity will make lagged health terms endogenous if ignored, and leads to new statistical models for econometric analysis in Markov chains. The results provide further confirmation that broadly-based preventive strategies may have an important role among the elderly. They have previously been hampered by the view that most modifiable risk factors have already had their full effects by the time individuals reach old age (Kaplan and Haan, 1989, p. 29).

The paper extends the applicability of Markovian modelling techniques by developing statistical models which are more suitable for multivariate analysis in health economics. This is discussed in Section 2. Section 3 sets out the econometric framework, and Section 4 discusses estimation methods. Section 5 describes the British Household Panel Survey data, and the results are presented in Section 6. Section 7 concludes the paper.

2. Markov chain models

Markov chain models, with and without explanatory variables, have been used in economics to analyse systems characterised by sequences of states taken on at discrete transition times (for a textbook treatment and comprehensive references, see Amemiya, 1985). At each transition time, there is either a change to a new state, or a renewal of the

[1] See Kaplan, Seeman, Cohen et al. (1987), Guralnik and Kaplan (1989), Lubben, Weiler and Chi (1989), Kaplan and Haan (1989), Guralnik et al. (1991), Roos and Havens (1991), Hollenbach et al. (1993), Jette, Feldman and Tennstedt (1993), Posner et al. (1993), Paganini-Hill and Hsu (1994), Burton et al. (1995), German et al. (1995), Mirand and Welte (1996), Stuck, Walthert et al. (1999).

state immediately before the transition. In *stationary first-order* models, the move to the next state is characterised by time-homogeneous transition probabilities reflecting the 'Markov property' i.e. that the past influences the future only through the present state.

The 'homogeneous' Markov chain model assumes that transition rates are the same for all individuals in a sample, and is often an inadequate approximation to data-generating processes. Its extensive use in previous studies may be partly due to practical difficulties surrounding the conventional 'logistic regression' approach developed by Boskin and Nold (1975) for multivariate variants of the Markov chain model. In health economics, applications may easily involve relatively large health-state spaces, and relatively large numbers of explanatory variables. In these cases, the logistic regression approach gives rise to many parameters and requires large sample sizes, since it involves the estimation of a distinct parameter vector for each independent transition probability. In addition, since many studies in health economics are based on self-assessed health data, the state space may be 'naturally ordered' (as in very poor, poor, fair, good and excellent self-assessed health). The logistic regression approach cannot distinguish between ordered and unordered multinomial responses, and may therefore ignore this feature of the data. Other common features, such as measurement errors and unobserved heterogeneity, can also not easily be addressed together in logistic regression models.

In this paper, new maximum likelihood methods are developed for multivariate Markov chains which exploit the ordinality of self-assessed health data to significantly reduce the number of parameters without losing any substantive information. Instead of estimating a separate parameter vector for each independent transition probability, an 'ordered probit' approach with thresholds is used to estimate a single parameter vector for each *row* of the transition matrix. This allows more variables to be included in the analysis than would otherwise be feasible, and also makes it relatively easy to control for self-reporting errors and time-persistent unobserved heterogeneity. As shown below, the latter would otherwise invalidate the assumption of a first-order Markov chain process, and also lead to the endogeneity of lagged health terms in a dynamic econometric model. Two ways of controlling for unobserved heterogeneity are explored in a 'random effects' set-up: numerical integration by Hermite quadrature, and a simulation approach which overcomes some potential difficulties with quadrature methods. The new models are applicable to a range of economic and social phenomena with naturally ordered states.

3. The econometric model

3.1. A dynamic demand-for-health function

The conceptual framework for this study is Grossman's (1972) health capital model. The representative consumer derives utility and income in each period $t = 1, 2, \ldots, T$ of the life-cycle from a stock of health capital H_t which depreciates exogenously with age at the rate $\delta_t \in (0, 1)$. The stock can be augmented by household production of gross health-investments using time and health-care as inputs. As the rate of depreciation increases with age, the stock of health falls below a minimum stock where death occurs. Decisions as to how much health to produce depend upon the utility function, relative prices of inputs, production possibilities, and depreciation rates. Length of life T is endogenous.

Most studies using the concept of health capital are based on two key equations of Grossman's model. First, health evolves according to a difference equation

$$H_t = (1 - \delta_{t-1})H_{t-1} + I_{t-1} \quad (1)$$

where I_{t-1} is gross health-investment during period t-1. Intertemporal utility maximisation subject to (1), a wealth constraint, and boundary conditions yields the second equation,

$$\tau_t + a_t = \{r + \delta_t - \tilde{\pi}_{t-1}\}\,\pi_t \quad (2)$$

This implicitly defines the *desired* stock of health capital at each t, denoted by H_t^* below. τ_t denotes the monetary value of the marginal utility of health, and a_t denotes the monetary value of the marginal product of health in producing 'healthy time'. The right-hand side is the marginal cost of capital, which is determined by the interest rate r (assumed to be constant for simplicity), the depreciation rate, the marginal cost of gross health-investment π_t, and its proportional rate of change, $\tilde{\pi}_{t-1}$.

Let the subscript n, n = 1, 2, . . ., N, denote the nth individual in a random sample from a large population. A suitable econometric model can be derived from (1) and (2) by interpreting them as the specifications of an implicit dynamic demand-for-health function linking H_{nt}, H_{nt-1}, and a vector $\mathbf{X}_{nt}$ of predetermined variables for each n. $\mathbf{X}_{nt}$ is assumed to explain the factors which determine the desired stock of health capital H_{nt}^* in equation (2), including the nth individual's tastes, rate of health deterioration, and efficiency of health production at each t. For example, Wagstaff (1993) assumes that H_{nt} evolves according to a Partial Adjustment Model (PAM) of the form

$$H_{nt} - H_{nt-1} = (1 - \alpha)(H_{nt}^* - H_{nt-1}) \quad (3)$$

where $\alpha \in (0, 1)$. The *desired* stock of health capital H_{nt}^* is modelled as

$$H_{nt}^* = \mathbf{X}_{nt}\tilde{\boldsymbol{\beta}} + \varepsilon_{nt} \quad (4)$$

where $\tilde{\boldsymbol{\beta}}$ is a coefficient vector, and ε_{nt} is a zero-mean error term reflecting the fact that $\mathbf{X}_{nt}$ cannot realistically contain all the variables influencing H_{nt}^*. Rearranging (3) gives

$$H_{nt} = \alpha H_{nt-1} + (1 - \alpha)H_{nt}^* \quad (5)$$

and substituting (4) into (5) yields a dynamic equation of the form

$$H_{nt} = \alpha H_{nt-1} + \mathbf{X}_{nt}\boldsymbol{\beta} + u_{nt} \quad (6)$$

where $\boldsymbol{\beta}$ is a vector of coefficients with the same signs as those in equation (4), and $u_{nt} = (1 - \alpha)\varepsilon_{nt}$ is a zero-mean disturbance derived from the disturbance term in (4).

In this reformulation of Grossman's model, equations (1) and (3) are complementary to one another. The PAM says that the *actual* net health-investment level at the beginning of a given period is a fraction $(1 - \alpha)$ of the level *desired* for that period. It is particularly

apt for older persons, since it is reasonable to assume that the elderly find it harder than the young to adjust health to optimal levels (e.g. via exercise, changes in habits etc.).[2]

It is assumed that all the variables on the right-hand-side of equation (6) are predetermined at time t i.e. $E[H_{nt-1}u_{nt}] = 0$ and $E[\mathbf{X}_{nt}u_{nt}] = \mathbf{0}$. However, this assumption is untenable when there is an unobserved individual component to ε_{nt} (and therefore u_{nt}) that is constant over time. This follows from the way in which ε_{nt} is introduced into the model, via equation (4). Since H_{nt-1}^* is determined by an equation analogous to (4) (with disturbance term ε_{nt-1}), and H_{nt-1} by an equation analogous to (5), a time-invariant individual component in ε_{nt-1} and ε_{nt} implies $E[H_{nt-1}u_{nt}] \neq 0$. The statistical model below attempts to control for this by allowing for a time-invariant unobserved component in ε_{nt}.

It should be noted that variables in $\mathbf{X}_{nt}$ may also be affected by time-invariant person-specific unobservable traits such as 'attitude to healthy living'. If these are correlated with a time-invariant component in u_{nt}, then it is again the case that $E[H_{nt-1}u_{nt}] \neq 0$ and $E[\mathbf{X}_{nt}u_{nt}] \neq \mathbf{0}$. There are ways of controlling for correlation between random effects and time-varying regressors involving averaging over time (Zabel, 1992), but these could not be implemented here. This should be borne in mind when interpreting results.

Equation (6) is modified in three ways for the purposes of this study. First, it is viewed as the restricted form of a richer model which also allows the marginal health-effects of the variables in $\mathbf{X}_{nt}$ to vary with the state of health at time t-1. For each n, let the sets A_j, j = 0, 1, 2, . . ., J, partition the state space S* of the stock of health capital so that, at each t, $H_{nt} \in A_j$ for some j = 0, 1, 2, . . ., J, $S^* = \bigcup_{j=0}^{J} A_j$, and $A_i \cap A_j = \varnothing$ for i ≠ j. Define a dummy variable y_{njt} as taking the value 1 if $H_{nt} \in A_j$, j = 0, 1, 2, . . ., J, and the value 0 otherwise. The dynamic demand-for-health function used in this paper is of the form

$$H_{nt} = \mathbf{X}_{nt}\boldsymbol{\beta} + \mathbf{X}_{nt}\boldsymbol{\alpha}_1 y_{n1t-1} + \mathbf{X}_{nt}\boldsymbol{\alpha}_2 y_{n2t-1} + \cdots + \mathbf{X}_{nt}\boldsymbol{\alpha}_J y_{nJt-1} + u_{nt} \qquad (7)$$

where each $\boldsymbol{\alpha}_j$, j = 1, 2, . . ., J, is a coefficient vector with the same dimension as $\boldsymbol{\beta}$. Equation (7) includes (6) as a special case if $\mathbf{X}_{nt}$ contains a term that is constant over n.

A second modification of equation (6) is required because the four-year observation period in this study is too short to allow significant variation in many of the variables in $\mathbf{X}_{nt}$. It is therefore assumed that each individual's intertemporal health-investment behaviour over a four-year period is predictable on the basis of a *single* vector $\mathbf{X}_n$ of explanatory variables at the *start* of the observation period. The variables in $\mathbf{X}_n$ reflect the usual arguments of demand-for-health functions derived from the health capital model. They include socioeconomic variables such as income and education, which are expected to have positive coefficients in equation (7), and determinants of health depreciation such as age, smoking behaviour and solitude, which are expected to have negative coefficients.

The third modification of equation (6) is the decomposition of u_{nt} into time-invariant and time-varying components, reflecting possible unobserved heterogeneity. Thus, let

$$u_{nt} = v_n + w_{nt} \qquad (8)$$

[2] Grossman's (1972) model assumes instantaneous adjustment to equilibrium. In the context of Wagstaff's empirical reformulation, this implies $\alpha = 0$ in equations (3), (5) and (6).

where $w_{nt} \sim N(0, \sigma_w^2)$ is a zero-mean disturbance term assumed to be serially uncorrelated and uncorrelated with all the regressors in equation (7), and $v_n \sim N(0, \sigma_v^2)$ is a time-invariant individual-specific 'random effect', assumed to be uncorrelated with w_{nt} and with all the variables in $\mathbf{X}_n$. The 'random effects' approach is used since it is not as easy to remove 'fixed effects' by conditioning on a minimal sufficient statistic in the context of the latent variable statistical model developed below (see Chamberlain, 1980).

3.2. Parameterising probabilities

Define the state space partition of S* to be the intervals $A_0 \equiv (-\infty, 0]$, $A_1 \equiv (0, c_1]$, $A_2 \equiv (c_1, c_2]$, . . ., $A_J \equiv (c_{J-1}, +\infty)$, where $-\infty < c_1 < c_2 < \cdots < c_{J-1} < +\infty$. For notational convenience, let $c_{0-1} \equiv -\infty$, $c_0 \equiv 0$, and $c_J \equiv +\infty$. The thresholds between states, c_j, j = 1, 2, . . ., J-1, are parameters (to be estimated) which are assumed to be the same for all individuals. They determine a mapping from the continuous state space S* to a discrete state space $S \equiv \{0, 1, \ldots, J\}$ of the form $A_j \mapsto j$ for j = 0, 1, 2, . . ., J. To account for measurement errors in self-assessments of health, it is assumed that the nth individual views the thresholds as c_{nj}^*, j = 1, 2, . . ., J-1, where

$$c_{nj}^* = c_j + z_{nj} \tag{9}$$

and where $z_{nj} \sim N(0, \sigma_z^2)$ denotes an individual-specific measurement error for all n and j which is assumed to be uncorrelated with the individual-specific random effect v_n in (8). Note that under the stochastic assumptions made about w_{nt} and z_{nj}, the random variable $(w_{nt} - z_{nj})$ is normal, with mean zero and variance $\tilde{\sigma}^2 \equiv \mathrm{Var}[w_{nt} - z_{nj}]$ for all n, j and t.

Given the above, and *conditional* on the unobserved heterogeneity v_n in (8), the nth individual's health follows a first-order stationary Markov chain with one absorbing state. Note that the first-order property is not true unconditionally of v_n, as it causes correlation between future health and all past health states even after controlling for present health.

The nth individual has a time-homogeneous transition probability matrix $\mathbf{P}_n$ with typical element P_{njk} in the jth row and kth column. P_{njk} denotes the conditional probability of a transition to state k at time t, given state j at time t-1 and the random effect v_n. The transition probabilities in each row of $\mathbf{P}_n$ satisfy the condition $\sum_k P_{njk}=1$. Assuming that 0 is the absorbing state in S, the transition matrix for the nth individual is

$$\mathbf{P}_n \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ P_{n10} & P_{n11} & P_{n12} & P_{n13} & \cdots & P_{n1J} \\ P_{n20} & P_{n21} & P_{n22} & P_{n23} & \cdots & P_{n2J} \\ P_{n30} & P_{n31} & P_{n32} & P_{n33} & \cdots & P_{n3J} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{nJ0} & P_{nJ1} & P_{nJ2} & P_{nJ3} & \cdots & P_{nJJ} \end{bmatrix} \tag{10}$$

Let $v_n^* \equiv v_n/\sigma_v \sim N(0, 1)$, $\rho^2 \equiv \sigma_v^2/(\sigma_v^2 + \tilde{\sigma}^2)$ and $\rho^* \equiv \rho/(1 - \rho^2)^{1/2}$. Then $\rho^* v_n^* \equiv v_n/\tilde{\sigma}$. Using (7), (8) and (9), the generic transition probability for $j \neq 0$ conditional on v_n is then

$$P_{njk} \equiv \mathrm{Prob}(H_{nt} \in A_k | H_{nt-1} \in A_j, v_n) = \mathbf{\Phi}(c_k^* - \mathbf{X}_n\boldsymbol{\beta}_j^* - \rho^* v_n^*) - \mathbf{\Phi}(c_{k-1}^* - \mathbf{X}_n\boldsymbol{\beta}_j^* - \rho^* v_n^*) \tag{11}$$

where Φ is the standard normal c.d.f., $c_k^* \equiv c_k/\tilde{\sigma}$, and $\beta_j^* \equiv (\beta + \alpha_j)/\tilde{\sigma}$. Note that $P_{njk} = 1$ for $j = k = 0$, and $P_{njk} = 0$ for $j = 0 \neq k$, since it is impossible to leave the absorbing state.

The estimation procedures described below provide estimates of the 'scaled' threshold parameters c_k^* and coefficient vectors β_j^*, and also an estimate of ρ^* which can be used to test for unobserved heterogeneity. Note that $\tilde{\sigma}$, β and α_j are individually unidentified. The source of identification of ρ^* in (11) is the fact that it is the only coefficient which is constant over all rows and columns of (10).

Given the stochastic assumptions of the model, equation (11) shows that the only effect of measurement errors is to alter the scaling parameter $\tilde{\sigma}$. The scaling parameter is arbitrary anyway, and only the magnitudes (not the signs or 't-ratios') of the estimates are affected. However, since many elderly people are retired from the workforce (approximately 70 per cent of the sample in this study), it may be difficult to measure socioeconomic status accurately. It is therefore necessary to consider the possibility that socioeconomic variables in $\mathbf{X}_n$ may also be subject to measurement errors which are correlated with measurement errors in self-assessed health. Under the reasonable assumptions that the measurement errors in $\mathbf{X}_n$ are additive and normally distributed, and that all error variances and covariances are independent of n, j, k, and t, the only effect of this would be to change the scaling parameter in the expression for P_{njk}, and make it depend on the row index j (via α_j in β_j^*). Since $\tilde{\sigma}$ is unidentified, the only *discernible* effect would be variation in the thresholds across the rows (as well as columns) of (10).

Equation (11) can be modified to allow for row-specific thresholds, but all attempts to implement such models in this paper gave rise to negative estimated probabilities, and failure of the estimation procedures below. This could be due to weak identification.

4. Maximum likelihood estimation and inference

4.1. Without unobserved heterogeneity

It is assumed that a random sample of N individuals is obtained from a large population at time 0. As well as observing an explanatory variable vector $\mathbf{X}_n$ for each individual, it is assumed that the state occupied in each period by each individual is observed for a specified number of equidistant time periods. The transition data required to estimate the models in this paper are then defined as $C(n, j, k) \equiv$ the number of times individual n makes a transition from state j in one period to state k in the next during the entire observation period, $j = 1, 2, \ldots, J$, $k = 0, 1, \ldots, J$.

The initial distribution of respondents across health states is taken as given, since the focus of this study is on the *dynamics* of health, and modelling the initial distribution using cross-sectional probability models adds nothing to the dynamic information provided by the transition data. Assuming a multinomial response model for self-assessed health, *and no unobserved heterogeneity*, the kernel of the log-likelihood function for the stationary first-order Markov chain model (denoted by $LL_{MC}(\theta)$) is

$$LL_{MC}(\theta) = \sum_n \log_e \left(\prod_j \prod_k (P_{njk})^{C(n, j, k)} \right) = \sum_n \sum_j \sum_k C(n, j, k) \log_e(P_{njk}) \qquad (12)$$

for n = 1, 2, . . ., N, j = 1, 2, . . ., J, and k = 0, 1, 2, . . ., J, where the P_{njk} are given by (11) with $v_n^* \equiv 0$, and where θ is a vector containing all the parameters to be estimated (cf. Boskin and Nold, 1975, Amemiya, 1985). Note that θ does *not* contain ρ^* here.

Maximum likelihood estimators for the parameters in θ are obtained by equating the first partials of $LL_{MC}(\theta)$ to zero, and solving them in terms of the transition data defined above. Since these partials are nonlinear, a numerical method of solution is required.

This paper uses the BFGS (Broyden, Fletcher, Goldfarb and Shanno) 'secant' algorithm in the MAXLIK application for GAUSS v. 386i. $LL_{MC}(\theta)$ is a well-behaved function of the unknown parameters, and the conditions are met for the maximum likelihood estimators to be Best Asymptotically Normal. Let $\hat{\theta}_N$ denote the maximum likelihood estimator (based on a sample of size N) of θ, and let LL_n denote the nth individual's log-likelihood. The estimated Hessian in the BFGS algorithm, i.e. $\sum_n[\partial^2 LL_n/\partial\hat{\theta}_N\,\partial\hat{\theta}_N{}']$, is -1 times the sample information matrix, whose inverse consistently estimates the asymptotic covariance matrix of $\hat{\theta}_N$ if the model is correct.

4.2. 'Integrating out' unobserved heterogeneity by Hermite quadrature

If there is unobserved heterogeneity, the random effects v_n^* must be accounted for, and the parameter ρ^* estimated. One way to do this is to take the expectation of each individual's likelihood function with respect to the (standard normal) density of v_n^*. This has the effect of 'integrating out' v_n^*. Letting $E_\phi[\cdot]$ denote expectation with respect to the standard normal p.d.f., the correct kernel of the log-likelihood (denoted by $LL_{MC}^*(\theta)$) is

$$LL_{MC}^*(\theta) = \sum_n \log_e \left(E_\phi \left[\prod_j \prod_k (P_{njk})^{C(n,j,k)} \right] \right)$$

$$= \sum_n \log_e \left(\int_{-\infty}^{\infty} \prod_j \prod_k (P_{njk})^{C(n,j,k)} f(v_n^*) dv_n^* \right) \qquad (13)$$

where the P_{njk} are given by (11), and where θ now contains the coefficient ρ^* of v_n^*. Testing the null hypothesis $\rho^* = 0$ provides a test for unobserved heterogeneity.

The integral in (13) can be approximated by Hermite quadrature (cf. Butler and Moffit, 1982). This involves constructing a set of likelihoods for each individual by substituting quadrature points for the (unobserved) random effects in (11), and then computing a weighted sum of these likelihoods for each individual. The (approximate) log-likelihood for an individual is the natural log of this weighted sum. For example, in ten-point quadrature, each individual's likelihood is evaluated at each of ten quadrature points, producing ten likelihoods. For each individual, a weighted sum of these ten likelihoods is computed, the weights corresponding to the points at which the likelihoods are evaluated. The natural log of this sum approximates the individual's log-likelihood.

Ten-point quadrature is used in this paper, in conjunction with the BFGS algorithm in the MAXLIK application for GAUSS v. 386i. Hermite quadrature points and weights are provided by Abramowitz and Stegun (1972, p. 924).

4.3. A simulation-based approach

Hermite quadrature can be problematic in the context of iterative maximum likelihood estimation because the integrands (which are likelihoods rather than *log*-likelihoods) can be extremely small. They can easily exceed the range of floating point numbers that digital computers can handle, and may then be perceived as being zero. The Hermite quadrature approach exacerbates this problem because the weights it uses are between 0 and 1, and also small. MAXLIK will 'crash' if an individual's log-likelihood cannot be computed in a particular iteration due to a near-zero argument in the natural log function. To confirm the results from the quadrature model in Section 4.2, this section proposes a Monte Carlo simulation approach to the approximation of (13) which also provides a useful alternative to Hermite quadrature if the integrand is too near zero.

The simulation approach is based on reconstructing the unobserved random effects in (11) by substituting i.i.d. random points for the fixed points in the Hermite quadrature. These random points are drawn from the standard normal distribution of the v_n^* in (11). Suppose M such random points are drawn, and let P^i_{njk} denote the transition probability P_{njk} in (11) evaluated at the ith random point. A Monte Carlo approximation of (13) is

$$\sum_n \log_e \left((1/M) \sum_i \left\{ \prod_j \prod_k (P^i_{njk})^{C(n,j,k)} \right\} \right) \tag{14}$$

The arguments of the natural logs in (14) are sample averages of the individuals' likelihoods evaluated at each of the M randomly drawn points. Basic stochastic limit theory shows that (14) converges almost surely to (13) as $M \to \infty$. Furthermore, the sample averages of the individuals' likelihoods evaluated at the M randomly drawn points converge in distribution to normal random variates as $M \to \infty$, with means equal to the corresponding expected individual likelihoods in (13), and standard deviations

$$\sigma_n^{mc} = \sqrt{(1/M) \mathrm{Var}\left[\prod_j \prod_k (P^i_{njk})^{C(n,j,k)} \right]} \tag{15}$$

The simulation approach will usually work, even when Hermite quadrature fails, because the factor (1/M) in (14) drops out of the maximum likelihood calculations (it is not a function of θ), leaving the kernel

$$LL_{MC}^{mc}(\theta) = \sum_n \log_e \left(\sum_i \left\{ \prod_j \prod_k (P^i_{njk})^{C(n,j,k)} \right\} \right) \tag{16}$$

The arguments of the natural logs in (16) can be many times larger than the corresponding arguments in Hermite quadrature, depending on the size of M.

A suitable sample size M can be calculated using *Chebyshev's Inequality* if the error tolerance is expressed as a multiple $k > 0$ of σ_n^{mc} in (15) (an extremely small quantity). The inequality says that the sample average of an individual's likelihoods evaluated at the M randomly drawn points will lie within $k\sigma_n^{mc}$ of the corresponding expected value in

(13) with a probability greater than or equal to $1 - (1/Mk^2)$. Thus, the sample average will lie within one standard deviation of its expected value in (13) with probability ≥ 90 per cent if M = 10, ≥ 95 per cent if M = 20, ≥ 98 per cent if M = 50, etc.

Parameter estimates are obtained by maximising (16) with respect to θ (which includes the coefficient ρ^* of the random effects). This is implemented in this paper using the BFGS algorithm in the MAXLIK application for GAUSS v. 386i.

5. The data

The analysis is based on individual-level data from the first five waves (1991-1995) of the British Household Panel Survey (BHPS) (Taylor, 1996). The BHPS is particularly suitable for this study because it has an extensive set of questions on health, and on the relevant social, socioeconomic and behavioural variables. It also includes relatively large numbers of elderly respondents. Another appealing feature of the BHPS is that losses due to attrition have been kept to a minimum, averaging approximately 5 per cent per annum for the survey as a whole over 1991-1995. Note that a 'balanced panel' is required for the models in this paper, although they could be extended to deal with censored observations.

After listwise deletion of missing cases, the samples consist of 785 men and 1022 women aged 60 and above in December 1991. The age of 60 was chosen as the cutoff point because it is by far the most commonly used criterion of 'old age' in the epidemiologic literature. A small number of elderly individuals who left the BHPS between 1991 and 1995 for reasons other than death are excluded from the study, although some may have left for health reasons (e.g. those who were institutionalised). There were too few such cases to warrant any special statistical adjustments.

Definitions and means for the explanatory variables are provided in Table 1. Indicators of health problems in 1991 that are likely to persist through time include an indicator of frequent General Practitioner visits (GP91), and a dummy for hospital inpatient stays (HOSP91). These are included along with AGEM (derived from AGE in Table 1) to control for differences in the dynamics of health not due to *modifiable* risk factors. Note that age (and income below) are dichotomised for the sake of consistency with the other variables in the study, which are all discrete. As a check, the basic models were re-estimated using the original variables, but there were no substantive differences in results.

Modifiable risk factors assumed to increase the rate of health depreciation are smoking behaviour (SMOKER), and solitude (proxied by SINGLE). A growing number of epidemiologic studies suggest that preventive efforts directed at smoking in older persons may result in decreased risk of morbidity and mortality (Kaplan and Haan, 1989, Hollenbach et al., 1993, Jette, Feldman and Tennstedt, 1993, Paganini-Hill and Hsu, 1994). Isolation/loneliness has also been found to significantly increase mortality risk in older age-groups (see e.g. Kaplan and Haan, 1989, p.38). It is frequently proxied by marital status variables in health economics (e.g. Wagstaff, 1986). Note that the variable SINGLE was chosen to reflect 'solitude', rather than other variables in the BHPS such as 'living alone', because the latter had low variability among the elderly: few elderly respondents in the BHPS actually live alone. In addition, it is conceivable that elderly individuals who live with friends or relatives, but who lack the close companionship of a partner, may experience 'loneliness' to an extent that affects their health (e.g. through depression), but this would be missed by variables such as 'living alone'.

Four indicators of socioeconomic position in 1991 are used in order to minimise measurement errors: a dummy variable (INCOMEM) indicating whether or not annual household income per person in the household exceeds the sample mean (derived from AHINCOME in Table 1), dummies for housing tenure (OWNHOUSE) and car ownership (ACARUSE), and an education dummy (COLLEGE). Education is treated as an indicator of socioeconomic position, although in Grossman's theory it may also have a efficiency role which is not separately identifiable here.

Table 2 shows the transition data used to estimate the models in this paper. The self-assessed health data come from variables in the BHPS with *five* health categories (very poor, poor, fair, good, excellent). The very poor and poor categories are combined into a single 'poor' category to avoid distortions due to small cell frequencies. The variables PO-DE to EX-EX count the number of transitions made between different pairs of health states in the four-year observation period. Thus, each individual can make at most 4 transitions between 'active' health states, while the variables PO-DE, FA-DE, GO-DE and EX-DE have a maximum value of 1. Note that numerous previous studies have found self-assessed health to be a 'reliable' health indicator (see e.g. van Doorslaer, 1987, and references therein), but, as a precautionary measure, the statistical model described in Section 3 is designed to accommodate the possibility of self-reporting errors.

6. Estimation results

Table 3 presents estimation results from the stationary first-order Markov chain model *without* controlling for unobserved heterogeneity. Tables 4 and 5 present random effects models using Hermite quadrature and the simulation-based approach (with M = 10) respectively. The simulation models were also estimated with M = 20, 25, 30, 35, 40, 45 and 50 to compare convergence times (Table 6), but as the coefficient estimates are all almost identical to those in Table 5, they are not presented here. All the models had the same set of starting values for the relevant parameters, which were obtained from 'pilot' runs of the GAUSS routine for the model in Table 3 with arbitrary starting values. The convergence tolerance for the log-likelihood gradients was set at 1×10^{-5} i.e. convergence occurred when all parameter gradients were zero to an accuracy of at least five decimals.

The state space definitions are 0 ≡ absorbing state for deaths, 1 ≡ poor health, 2 ≡ fair health, 3 ≡ good health, 4 ≡ excellent health. Each column in Tables 3 to 5 relates to one *row* of the transition matrix (10) described in Section 3, and presents parameter estimates for the corresponding values of j and k in equation (11).

All the results presented here are based on combined data for men and women. A dummy variable (SEX) is included in the specifications to control for differences by sex, taking the value 1 for men, and 0 for women (mean = 0.43). Similar results were obtained for men and women separately, so these are not presented here.

The coefficients of SEX in Table 3 are all negative, and highly significant in Rows 1 through 4. This suggests that men are less likely to make transitions to the better health states than women, ceteris paribus. The indicators of health problems in 1991 also have highly significant negative coefficients in all the rows of the transition matrix. Not surprisingly, this suggests that elderly individuals who have health problems are also less likely to make subsequent transitions to the better health states, ceteris paribus.

Among the health depreciation variables, the coefficients of the age dummy are negative and highly significant in all the rows of the transition matrix in Table 3, confirming that older individuals are less likely to make transitions to the better health states, ceteris paribus. SMOKER has a highly significant negative impact on health in the fair, good and excellent health states, reinforcing the epidemiologic evidence cited earlier. Its coefficient in the poor health state is also negative, but not significant.

SINGLE has a negative impact on health in the fair, good and excellent health states, which seems to confirm the epidemiologic evidence (cited earlier) suggesting that loneliness among the elderly increases the probability of transitions to inferior health states. It is not statistically significant in Row 4. It has a positive coefficient in the poor/very poor health state, but this is only just significant, and not significant in Table 4.

The socioeconomic variables INCOMEM, ACARUSE and COLLEGE have highly significant positive coefficients in one or more of Rows 2 through 4 of the transition matrix in Table 3, suggesting that socioeconomic status remains important as a target variable for preventive strategies at old age. Only OWNHOUSE fails to achieve statistical significance in all three rows after controlling for the other socioeconomic variables. Negative coefficients for INCOMEM and COLLEGE in Row 1 contrast with highly significant positive coefficients for OWNHOUSE and ACARUSE, and suggest that the role of socioeconomic status is more ambiguous among individuals in the poor/very poor health state.

The results in Tables 4 and 5 are substantively similar to those in Table 3, although there are some minor differences in the magnitudes of the estimates. They seem to confirm that smoking, loneliness and socioeconomic status remain significant risk factors at old age even after *simultaneously* controlling for past health, measurement errors and unobserved heterogeneity. To the author's knowledge, no previous economic or epidemiologic study in this area has explicitly controlled for these together.

In all cases, the estimated random effect parameter fails to achieve statistical significance, indicating that unobserved heterogeneity is not significant in this data set. The simulation approach used to obtain the results in Table 5 worked very well, and seems to provide a useful alternative to Hermite quadrature. The model with $M = 10$ converged in only 50 iterations (455 minutes), compared to 78 iterations (724 minutes) for the ten-point Hermite quadrature. Convergence times and maximised values of the log-likelihood for different values of M are given in Table 6. The number of iterations before convergence continues to fall as M rises, stabalising at around 40 iterations for $M \geq 25$. However, the time per iteration rises, making the total run time higher for higher values of M. Note that parameter estimates for all the higher values of M were almost identical to those in Table 5, suggesting that there was little gain in raising M above 10.

7. Summary and concluding remarks

The 'ordered probit' approach to Markov chain modelling in this paper has allowed larger state spaces, and larger numbers of variables to be analysed than is computationally feasible using conventional unordered logistic regression models. The approach can be applied whenever state spaces are 'naturally' ordered, without losing any substantive information. Note that an 'ordered logit' model could be used instead of the ordered probit approach, although the normality assumption underlying the probit

approach makes it easier to model the effects of measurement errors and unobserved heterogeneity. A simulation approach to dealing with unobserved heterogeneity was also developed in the paper, and found to compare favourably with Hermite quadrature.

The new techniques were used to estimate dynamic multivariate models of health status transitions in samples of elderly men and women from the 1991-1995 waves of the British Household Panel Survey. Modifiable risk factors such as solitude, low socioeconomic status, and smoking were found to be significant predictors of transitions from better to worse health states, even after controlling simultaneously for past health, measurement errors and unobserved heterogeneity. This seems to suggest that broadly-based preventive strategies aimed at social, socioeconomic and behavioural risk factors may have an important role to play in the non-institutionalised elderly population.

The findings from this study tentatively suggest that an 'environmental' approach to disease prevention and health-promotion (see e.g. Syme, 1986), in combination with more traditional individually-oriented approaches, may be effective among the elderly. For example, steps to improve the safety of elderly people may increase their willingnes to go outside, decrease their social isolation, and lead to better health. In addition, many features characteristic of the social and physical environments lived in by people in the lower socioeconomic strata may contribute to poor health. Interventions aimed at improving these environments may help to reduce the health effects of low socioeconomic status among older persons. National anti-smoking campaigns, usually aggressively aimed at the young, may also be worth directing at the elderly population. The findings from this study by no means 'prove' that such schemes will work, but support the view that they should at least be considered by policy makers.

A question arises as to the desirability of health-promotion schemes that prolong life among the elderly, if the end result is simply that individuals experience morbidity for longer, and that the demand for health-care and the cost to the taxpayer are greater than they would otherwise be. In this regard, some authors (e.g. Fries, 1984) have argued that prevention at old age will not lead to longer life-spans, but to 'compression of morbidity' to shorter periods just before death. Thus, prevention could reduce health-care demand.

References

Amemiya, T., 1985, Advanced econometrics (Cambridge, MA: Harvard University Press).

Boskin, M., and F. Nold, 1975, A Markov model of turnover in aid to families with dependent children, Journal of Human Resources 10, 476-481.

Burton, L., D. Steinwachs, P. German, S. Shapiro, L. Brant, T. Richards, and R. Clark, 1995, Preventive services for the elderly: would coverage affect utilization and costs under Medicare?, American Journal of Public Health 85, 387-391.

Butler, J., and R. Moffitt, 1982, A computationally efficient quadrature procedure for the one factor multinomial probit model, Econometrica 50, 761-764.

Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of Economic Studies 47, 225-238.

Fries, J., 1984, The compression of morbidity: miscellaneous comments, Gerontologist 24, 354-359.

German, P., L. Burton, S. Shapiro, D. Steinwachs, I. Tsuji, M. Paglia, and A. Damiano, 1995, Extended coverage for preventive services for the elderly: response and results in a demonstration population, American Journal of Public Health 85, 379-386.

Grossman, M., 1972, On the concept of health capital and the demand for health, Journal of Political Economy 80, 223-255.

Guralnik, J., and G. Kaplan, 1989, Predictors of healthy aging: prospective evidence from the

Alameda County Study, American Journal of Public Health 79, 703-708.
Guralnik, J., A. Lacroix, L. Branch, S. Kasl, and R. Wallace, 1991, Morbidity and disability in older persons in the years prior to death, American Journal of Public Health 81, 443-447.
Haynes, R., 1991, Inequalities in health and health service use: evidence from the General Household Survey, Social Science and Medicine 33, 361-368.
Hollenbach, K., E. Barrett-Connor, S. Edelstein, and T. Holbrook, 1993, Cigarette smoking and bone mineral density in older men and women, American Journal of Public Health 83, 1265-1270.
Jette, A., H. Feldman, and S. Tennstedt, 1993, Tobacco use: A modifiable risk factor for dental disease among the elderly, American Journal of Public Health 83, 1271-1276.
Kaplan, G., T. Seeman, R. Cohen, L. Knudsen, and J. Guralnik, 1987, Mortality among the elderly in the Alameda County Study: Behavioural and demographic risk factors, American Journal of Public Health 77, 307-312.
Kaplan, G., and M. Haan, 1989, Is there a role for prevention among the elderly ? Epidemiological evidence from the Alameda County Study, in: M. G. Ory and K. Bond, eds. Aging and health care: Social science and policy perspectives (Routledge) 27-51.
Lubben, J., G. Weiler, and I. Chi, 1989, Health practices of the elderly poor, American Journal of Public Health 79, 731-734.
Mirand, A., and J. Welte, 1996, Alcohol consumption among the elderly in a general population, Erie County, New York, American Journal of Public Health 86, 978-984.
Norton, E., 1992, Incentive regulation of nursing homes, Journal of Health Economics 11, 105-28.
Paganini-Hill, A., and G. Hsu, 1994, Smoking and mortality among residents of a California retirement community, American Journal of Public Health 84, 992-995.
Portrait, F., M. Lindeboom, and D. Deeg, 1999, Health and mortality of the elderly: the grade of membership method, classification and determination, Health Economics 8, 441-457.
Posner, B., A. Jette, K. Smith, and D. Miller, 1993, Nutrition and health risks in the elderly: The nutrition screening initiative, American Journal of Public Health 83, 972-978.
Roos, N., and B. Havens, 1991, Predictors of successful aging: a twelve year study of Manitoba elderly, American Journal of Public Health 81, 63-68.
Stuck, A., J. Walthert, T. Nikolaus, C. Bula, C. Hohmann, and J. Beck, 1999, Risk factors for functional status decline in community-living elderly people: a systematic literature review, Social Science & Medicine 48, 445-469.
Syme, S., 1986, Strategies for health promotion, Preventive Medicine 16, 492-507.
Taylor, M., ed., 1996, British Household Panel Survey User Manual, Volumes A and B, Essex University, Institute for Social and Economic Research.
Van Doorslaer, E., 1987, Health, knowledge and the demand for medical care (van Gorcum, Wolfeboro, NH).
Wagstaff, A., 1986, The demand for health: some new empirical evidence, Journal of Health Economics 5, 195-233.
Wagstaff, A., 1993, The demand for health: An empirical reformulation of the Grossman model, Health Economics 2, 189-198.
Zabel, J., 1992, Estimating fixed and random effects models with selectivity, Economics Letters 40, 269-272.

Table 1. Explanatory variables: definitions and means

Variable	Definition	Mean Men N= 785	Mean Women N= 1022
Health problems			
GP91	=1 if at the time of the interview in 1991, the respondent had consulted a GP or family doctor about his/her own health more than once or twice since 1.9.90	0.45	0.52
HOSP91	=1 if at the time of the interview in 1991, the respondent had been in a hospital or clinic as an inpatient, overnight or longer, since 1.9.90	0.13	0.13
Health depreciation			
AGE	Age in years on 1.12.91	69.75	70.64
AGEM	=1 if AGE exceeds the sample mean	0.47	0.45
SMOKER	=1 if the respondent smoked cigarettes in 1991	0.23	0.22
SINGLE	=1 if neither married nor living with a partner in 1991	0.27	0.56
Socioeconomic status			
AHINCOME	Annual household income (1.9.90-1.9.91) in thousands of pounds per person in household (includes imputed values)	5.91	5.33
INCOMEM	=1 if AHINCOME exceeds the sample mean	0.36	0.35
OWNHOUSE	=1 accommodation was either owned by the household or on mortgage in 1991	0.69	0.61
ACARUSE	=1 if owned or had use of a car or van in 1991	0.63	0.27
COLLEGE	=1 if the respondent had a vocational or academic college education (includes university degrees, teaching, nursing and other higher qualifications)	0.21	0.15

Table 2. Transition data: definitions and means

Variable	Definition	Mean Men N= 785	Mean Women N= 1022
PO-DE	=1 if respondent died while in poor health state	0.06	0.04
PO-PO	number of transitions from poor to poor between 1991 and 1995	0.14	0.16
PO-FA	number of transitions from poor to fair between 1991 and 1995	0.12	0.13
PO-GO	number of transitions from poor to good between 1991 and 1995	0.03	0.05
PO-EX	number of transitions from poor to excellent between 1991 and 1995	0.002	0.01
FA-DE	=1 if respondent died while in fair health state	0.06	0.04
FA-PO	number of transitions from fair to poor between 1991 and 1995	0.11	0.12
FA-FA	number of transitions from fair to fair between 1991 and 1995	0.43	0.55
FA-GO	number of transitions from fair to good between 1991 and 1995	0.26	0.27
FA-EX	number of transitions from fair to excellent between 1991 and 1995	0.01	0.02
GO-DE	=1 if respondent died while in good health state	0.05	0.04
GO-PO	number of transitions from good to poor between 1991 and 1995	0.04	0.05
GO-FA	number of transitions from good to fair between 1991 and 1995	0.29	0.33
GO-GO	number of transitions from good to good between 1991 and 1995	1.02	1.14
GO-EX	number of transitions from good to excellent between 1991 and 1995	0.22	0.17
EX-DE	=1 if respondent died while in excellent health state	0.01	0.01
EX-PO	number of transitions from excellent to poor between 1991 and 1995	0.01	0.005
EX-FA	number of transitions from excellent to fair between 1991 and 1995	0.04	0.03
EX-GO	number of transitions from excellent to good between 1991 and 1995	0.29	0.22
EX-EX	number of transitions from excellent to excellent between 1991 and 1995	0.38	0.31

Table 3. Estimation results without controlling for unobserved heterogeneity

	Rows of the transition matrix $\mathbf{P}_n$ *(equation (10))*			
	Row 1 (poor/v.poor)	Row 2 (fair)	Row 3 (good)	Row 4 (excellent)
Variable	(P_{n10} to P_{n14})	(P_{n20} to P_{n24})	(P_{n30} to P_{n34})	(P_{n40} to P_{n44})
constant	1.074***	1.714***	2.540***	3.534***
	(10.227)	(89.817)	(78.087)	(37.739)
SEX	-0.287***	-0.135***	-0.053***	-0.154***
	(-6.994)	(-6.763)	(-3.351)	(-2.966)
Health problems				
GP91	-0.182***	-0.181***	-0.298***	-0.405***
	(-3.180)	(-8.561)	(-9.492)	(-16.877)
AHOSP91	-0.258***	-0.140***	-0.193***	-0.055**
	(-9.811)	(-4.939)	(-12.335)	(-2.479)
Health depreciation				
AGEM	-0.181***	-0.106***	-0.116***	-0.184***
	(-5.894)	(-3.657)	(-4.679)	(-5.832)
SMOKER	-0.064	-0.054***	-0.167***	-0.113**
	(-1.305)	(-2.870)	(-6.746)	(-2.221)
SINGLE	0.075*	-0.063**	-0.056*	-0.051
	(1.693)	(-2.192)	(-1.813)	(-1.105)
Socioeconomic status				
INCOMEM	-0.066***	-0.038	0.163***	0.063
	(-3.532)	(-1.612)	(6.233)	(1.410)
OWNHOUSE	0.148***	-0.006	0.017	0.067
	(4.146)	(-0.298)	(0.683)	(1.221)
ACARUSE	0.159***	0.118***	0.192***	0.228***
	(8.859)	(5.992)	(8.888)	(4.970)
COLLEGE	-0.206***	0.256***	0.105***	0.210***
	(-2.918)	(15.387)	(3.846)	(5.399)
Thresholds				
c_1	0.708***	0.708***	0.708***	0.708***
	(34.223)	(34.223)	(34.223)	(34.223)
c_2	1.860***	1.860***	1.860***	1.860***
	(76.953)	(76.953)	(76.953)	(76.953)
c_3	3.607***	3.607***	3.607***	3.607***
	(123.884)	(123.884)	(123.884)	(123.884)

Maximised value of the log-likelihood=-7277.0239. Numbers in parentheses are t-ratios. *** Significant at the 1 per cent level in a two-tailed test (critical value = 2.5758). ** Significant at the 5 per cent level in a two-tailed test (critical value = 1.9600). * Significant at the 10 per cent level in a two-tailed test (critical value = 1.6449).

Table 4. Estimation results using 10-point Hermite quadrature to control for random effects

	Rows of the transition matrix $\mathbf{P}_n$ *(equation (10))*			
	Row 1 (poor/v.poor)	Row 2 (fair)	Row 3 (good)	Row 4 (excellent)
Variable	(P_{n10} to P_{n14})	(P_{n20} to P_{n24})	(P_{n30} to P_{n34})	(P_{n40} to P_{n44})
constant	1.073***	1.714***	2.540***	3.534***
	(8.458)	(20.506)	(31.579)	(32.802)
SEX	-0.287***	-0.135***	-0.053**	-0.154***
	(-7.027)	(-6.101)	(-1.978)	(-2.893)
Health problems				
GP91	-0.185***	-0.182***	-0.298***	-0.405***
	(-3.066)	(-6.837)	(-8.576)	(-17.494)
AHOSP91	-0.258***	-0.140***	-0.193***	-0.055**
	(-8.588)	(-4.797)	(-12.382)	(-2.376)
Health depreciation				
AGEM	-0.180***	-0.106***	-0.116***	-0.184***
	(-5.360)	(-3.697)	(-3.648)	(-5.828)
SMOKER	-0.063	-0.054***	-0.167***	-0.112**
	(-1.508)	(-2.926)	(-6.368)	(-2.252)
SINGLE	0.075	-0.063**	-0.056*	-0.051
	(1.611)	(-2.212)	(-1.902)	(-1.140)
Socioeconomic status				
INCOMEM	-0.066***	-0.038	0.163***	0.063
	(-2.871)	(-1.514)	(7.213)	(1.516)
OWNHOUSE	0.146***	-0.006	0.017	0.067
	(3.788)	(-0.278)	(0.614)	(1.178)
ACARUSE	0.159***	0.118***	0.192***	0.228***
	(6.732)	(5.779)	(8.423)	(5.572)
COLLEGE	-0.205***	0.256***	0.105***	0.210***
	(-3.629)	(13.521)	(3.303)	(4.898)
Thresholds				
c_1	0.708***	0.708***	0.708***	0.708***
	(32.915)	(32.915)	(32.915)	(32.915)
c_2	1.860***	1.860***	1.860***	1.860***
	(76.532)	(76.532)	(76.532)	(76.532)
c_3	3.607***	3.607***	3.607***	3.607***
	(121.513)	(121.513)	(121.513)	(121.513)
Random effects				
ρ*	0.0002	0.0002	0.0002	0.0002
	(0.0020)	(0.0020)	(0.0020)	(0.0020)

Maximised value of the log-likelihood=-7277.0239. Numbers in parentheses are t-ratios. *** Significant at the 1 per cent level in a two-tailed test (critical value = 2.5758). ** Significant at the 5 per cent level in a two-tailed test (critical value = 1.9600). * Significant at the 10 per cent level in a two-tailed test (critical value = 1.6449).

Table 5. Estimation results using a simulation-based approach to control for random effects (M = 10)

	Rows of the transition matrix $\mathbf{P}_n$ *(equation (10))*			
	Row 1 (poor/v.poor)	Row 2 (fair)	Row 3 (good)	Row 4 (excellent)
Variable	(P_{n10} to P_{n14})	(P_{n20} to P_{n24})	(P_{n30} to P_{n34})	(P_{n40} to P_{n44})
constant	1.074***	1.714***	2.540***	3.534***
	(10.150)	(69.779)	(74.921)	(36.716)
SEX	-0.287***	-0.135***	-0.053***	-0.155***
	(-6.832)	(-5.736)	(-2.764)	(-2.921)
Health problems				
GP91	-0.182***	-0.181***	-0.298***	-0.405***
	(-3.183)	(-7.723)	(-11.043)	(-14.933)
AHOSP91	-0.258***	-0.140***	-0.193***	-0.055**
	(-7.569)	(-4.289)	(-9.524)	(-2.127)
Health depreciation				
AGEM	-0.181***	-0.106***	-0.116***	-0.184***
	(-7.798)	(-3.368)	(-5.390)	(-5.490)
SMOKER	-0.064	-0.054**	-0.167***	-0.113**
	(-1.167)	(-2.396)	(-6.784)	(-2.164)
SINGLE	0.075*	-0.063**	-0.056*	-0.051
	(1.671)	(-2.222)	(-1.758)	(-1.056)
Socioeconomic status				
INCOMEM	-0.066***	-0.038	0.163***	0.063
	(-2.736)	(-1.450)	(6.219)	(1.388)
OWNHOUSE	0.146***	-0.006	0.017	0.067
	(3.991)	(-0.255)	(0.652)	(1.209)
ACARUSE	0.159***	0.118***	0.192***	0.228***
	(6.850)	(5.268)	(8.617)	(4.617)
COLLEGE	-0.206***	0.256***	0.105***	0.210***
	(-2.791)	(12.008)	(3.444)	(5.217)
Thresholds				
c_1	0.708***	0.708***	0.708***	0.708***
	(31.347)	(31.347)	(31.347)	(31.347)
c_2	1.860***	1.860***	1.860***	1.860***
	(71.131)	(71.131)	(71.131)	(71.131)
c_3	3.607***	3.607***	3.607***	3.607***
	(111.785)	(111.785)	(111.785)	(111.785)
Random effects				
ρ*	-0.0000	-0.0000	-0.0000	-0.0000
	(-0.0010)	(-0.0010)	(-0.0010)	(-0.0010)

Maximised value of the log-likelihood=-3116.2438. Numbers in parentheses are t-ratios. *** Significant at the 1 per cent level in a two-tailed test (critical value = 2.5758). ** Significant at the 5 per cent level in a two-tailed test (critical value = 1.9600). * Significant at the 10 per cent level in a two-tailed test (critical value = 1.6449).

Table 6. Convergence times for different values of the Monte Carlo sample size (M)

Monte Carlo sample size (M)	Maximised value of LL_{MC}^{mc} (equation (16))	Number of iterations	Convergence time (minutes)
10	-3116.24378	50	455.28650
20	-1863.73980	44	782.13817
25	-1460.51317	40	873.67883
30	-1131.05912	40	1044.96033
35	-852.50827	40	1243.59750
40	-611.21594	43	1514.12750
45	-398.38206	43	1682.84300
50	-207.99654	40	1730.90283