THE UNIVERSITY of York

Discussion Papers in Economics

No. 2001/19

Socio-Economic Status, Health and Lifestyle

by

Paul Contoyannis and Andrew M Jones

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Socio-economic status, health and lifestyle

Paul Contoyannis and Andrew M. Jones,[†]
Department of Economics and Related Studies,
University of York

December 11, 2001

Abstract

The role of lifestyle in mediating the relationship between socio-economic characteristics and health has been discussed extensively in the epidemiological and economic literatures. Previous analyses have not considered a formal framework incorporating unobservable heterogeneity. In this paper we develop a simple economic model in which health is determined(partially) by lifestyle, which depends on preferences, budget and time constraints and unobservable characteristics. We estimate a recursive empirical specification consisting of a health production function and reduced forms for the lifestyle equations using Maximum Simulated Likelihood for a multivariate probit model with discrete indicators of lifestyle choices and self-assessed health(SAH) on British panel data from the 1984 and 1991 *Health and Lifestyle Survey*. We find that prudent drinking and not smoking in 1984 have dramatic positive effects on the probability of reporting excellent or good SAH in 1991. The failure of epidemiological analyses to account for unobserved heterogeneity can explain their low estimates of the relevance of lifestyle in the socio-economic status-health relationship. Accounting for unobserved heterogeneity also leads us to conclude that indicators for sleep, exercise, and breakfast in 1984 are unimportant for SAH in 1991.

JEL codes I1 C0

Keywords: Determinants of health, lifestyles, simulation-based inference, panel data

[*]The authors wish to thank participants in the York Seminars in Health Econometrics. Data from the *Health and Lifestyle Survey*(HALS) were supplied by the ESRC Data Archive. Neither the original collectors of the data nor the archive bear any responsibility for the analysis or interpretation presented here. We are grateful for research funding from the ESRC (award no R000238169).

[†]Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, United Kingdom. Tel: +44-1904-433790; Fax: +44-1904-433759, Email: amj1@york.ac.uk.

1 Introduction

Analysts of the causes of inequalities in health have long recognised that variation in medical care utilisation cannot fully explain observed health differences (e.g. Auster et al.(1969), Evans et al.(1994)). A common feature of this literature has been the growing use of the concept of lifestyle by epidemiologists, sociologists, and economists, in order to categorise behavioural patterns and explain observed health inequalities. In an influential work, Fuchs (1986) argues that beyond a fairly low level in the provision of food, hygiene and basic health care, it is personal lifestyle that causes the greatest variation in health. McGinnis and Foege(1993) estimated that the three leading external (nondegenerative or directly genetically determined) causes of mortality in the U.S. in 1990 were tobacco, diet and activity, and alcohol consumption. They estimated that these lifestyle variables explained around 38% of premature mortality, and also noted that a dramatically reduced quality of life is associated with many of the diseases related to these behaviours. Other authors have concluded that, with exception of tobacco consumption, lifestyle factors do not affect the widely observed relationship between socio-economic status and health substantially (Borg and Kristensen(2000), Lantz *et al.*(1998), Power *et al.*(1998), Marmot *et al.*(1996), Lynch *et al.*(1996).)

Although the focus of research into the determinants of health has shifted, it is not easy to define lifestyle both comprehensively and empirically. The World Health Organisation (1986) consider a number of meanings, adopting a broad definition; '...the term 'lifestyle' is taken to mean a general way of living based on the interplay between living conditions in the wide sense and individual patterns of behaviour as determined by sociocultural factors and personal characteristics'. In this paper we adopt a narrow(and operationalizable) definition of lifestyle which focuses on health related behaviour and accords with the epidemiological literature on the determinants of health(e.g Lynch *et al.*(1996,1997), Marmot *et al.*(1997)). We define a lifestyle as a set of behaviours which are considered to influence health *a priori* and are generally considered to involve a considerable amount of free choice. In using this definition, there is no implication that other characteristics of an individual's environment, both natural and social, are inconsequential. We adopt an economic approach which recognises that individuals are making decisions that reflect the constraints of their circumstances, as well as their preferences. We develop a static model to identify interactions between health related behaviour and self-assessed health status, given other factors that are observable, such as socio-economic status, and unobservable heterogeneity.

Descriptive analysis of our data suggests a number of interesting features. Firstly,

based on the baseline assumption that lifestyle choices are independent and that all individuals have the same probabilities of each lifestyle choice, we can compare the expected number of individuals behaving completely 'healthily' or 'unhealthily' relative to the observed number. The expected values are around half the number of observed values. This suggests that health-related behaviours are not randomly distributed, but rather that healthy and unhealthy behaviours cluster together in certain individuals. This clustering may be due to observed or unobserved factors. Secondly, average self-assessed health gradually increases as the number of healthy behaviours increases. Thirdly, some of the variation in lifestyle choices appears to be related to observed characteristics of individuals. For example, the proportions of individuals in higher social groups gradually increase as we move from completely unhealthy to completely healthy lifestyles. Conversely, the proportions in lower social class groups gradually decrease. However, while these observations are indicative, they are correlations only; the use of appropriate econometric techniques offers the opportunity to assess questions of causality.

We estimate the structural parameters of a health production function, together with the reduced form parameters for the lifestyle equations using panel data from the *Health and Lifestyle Survey*(HALS) conducted in the United Kingdom in 1984 and 1991. This is achieved using Maximum Simulated Likelihood for a multivariate probit model with discrete indicators of lifestyle choices and self-assessed health(SAH). In addition to the substantive empirical results, this paper demonstrates the applicability and computational feasibility of models with flexible heterogeneity structures in the presence of multiple discrete outcomes in health economics.

The structure of the paper is as follows. Section 2 surveys previous economic literature in this area. Section 3 explains the theoretical model which forms the basis of our empirical analyses. Section 4 presents the U.K *Health and Lifestyle Survey*(HALS) dataset. Section 5 describes our estimation strategy, while the empirical results are discussed in section 6. Section 7 contains a short conclusion.

2 Previous economic literature

An example of the generic approach to the estimation of health production and input equations is Rozensweig and Schultz(1983). Rosenzweig and Schultz used instrumental variable techniques to examine the effect of health inputs on birth weight in the presence of unobservable heterogeneity. Their concern was to obtain consistent estimates of the parameters of the child health(birth weight) production function, while recognizing the difficulties created by input choices being influenced by unobservables that also influence

health outcomes. While a seminal paper in the health production literature, the results of Rosenzweig and Schultz indicate that the instruments employed in their two stage estimation approach had little explanatory power. An alternative, but less ambitious approach is to estimate reduced form health equations. This approach allows estimates of policy relevant parameters and predictions of the levels of health, given the levels of exogenous covariates, to be obtained. However, this strategy cannot identify the technological parameters of the health production function.

Kenkel(1995) estimated health production functions using several output measures, in order to estimate the impact of lifestyles on adult health. He modelled current health as dependent on previous health and the depreciation rate, as well as lifestyle and schooling. However, the empirical specifications were not derived from an explicit structural model. In addition, Kenkel's attempts at accounting for endogeneity were unsuccessful. As he notes, this was probably due to the lack of explanatory power of the first stage instruments; money prices are largely irrelevant for many elements of a lifestyle. The approach we adopt is able to exploit the panel nature of the data, while Kenkel(1995) was forced to rely on *OLS* estimates using cross-sectional US data from the 1985 National Health Interview Survey. Kenkel(1995) found excessive weight, cigarette smoking, heavy drinking, excessive or insufficient sleep, and stress to be harmful inputs in the health production function. Exercise and moderate alcohol consumption emerge as beneficial health inputs.[1]

Our approach also differs from work which has either focused on single behaviours such as smoking or drinking, and how they are determined with health (see e.g., Blaylock and Blisard(1992), Mullahy and Portney(1990)) or have examined interactions between lifestyle choices without the basis of an underlying structural model (see e.g., Hu et al. (1995)).

3 A simple model of lifestyle and health production

Becker's (1965) seminal work on the allocation of time provides our starting point. He focuses on the distinction between technology and preference orderings in the production and consumption of fundamental commodities. In our model the fundamental commodity adult health is produced by health related behaviours and other inputs, and also provides consumption benefits.[2] It is assumed that each health-related behaviour takes time to

[1] There are many other studies that have estimated the impact of endogenous treatments, but few have considered multiple factors simultaneously and a formal consideration of sources of unobserved heterogeneity.

[2] The term fundamental commodity does not indicate a commodity which is necessary to flourish or enjoy life. Rather, it denotes an argument of the direct utility function of an individual, as introduced by

consume(Gravelle and Rees (1992) pp157-164). Our model is characterized by joint production in that a subset of goods are inputs into the health production function, and in combination with a time input, they produce utility directly (Pollak and Wachter(1975), Grossman (1971), Rosenzweig and Schultz (1983)). Further, in order to provide a deterministic solution to the model, the individual is assumed to know, or believe that they know, the marginal productivity of health related behaviours and other parameters of their decision problem. Individuals are also assumed to know the health level produced by a given combination of inputs. However, some of these inputs will be unobservable to the researcher but known to the decision maker and will influence the levels of *desired* consumption. This endogeneity problem is considered further in section 5.

Income is assumed to be endogenous, but there is no direct influence of lifestyle or health on wages. This differs from the work of Suen and Mo (1994) who examine the concept of 'productive consumption.' In this simple model, the only benefit of health is utility, unlike Grossman(1972) and others (e.g. Dardanoni and Wagstaff(1987), Birch and Stoddart (1989), Ehrlich and Chuma(1990), and Forster(2001)), who consider dynamic models which allow for flows of future pecuniary and non-pecuniary benefits resulting from investment in the stock of health.

The utility maximization problem is:

$$\max_{C,H} U(C, H; X_U) \tag{1}$$

where U is a utility function to be maximized subject to budget and time constraints and a health production function. C represents an M-vector of goods, H represents a scalar measure of the individual's health, and X_U is a vector of exogenous influences on U. The health measure can be envisaged as current health or future health, recognizing that the influence of health-related behaviour on health is not immediate.

The health production function is:

$$H = h(C, X_H), \tag{2}$$

where C is as defined above and X_H is a vector of exogenous influences on health. For parsimony we combine X_U and X_H into the vector X. It is likely that health and utility will also be dependent on unobservable components such as genetic endowment. Hence

Lancaster (1966) and Becker (1965).

the vector X may include unobservable variables as well as observable variables. These will play an important role in the development of the empirical specifications and analyses. Note that, in general, not all the M goods will affect health, H.

The money budget constraint is:

$$\sum_{j=1}^{M} p_j C_j \leq I = m_0 + wL \ , \tag{3}$$

where m_0 is exogenous income and wL is labour income derived from L hours of labour at the wage rate w, which is assumed exogenous.

The time constraint is:

$$\sum_{j=1}^{M} \tau_j C_j = T - L \ , \tag{4}$$

where τ_j is the amount of time necessary to consume a unit of C_j. T denotes the total available time in a given period which is to be apportioned between labour and the consumption of goods. It is assumed that 'leisure' is used in the consumption of goods, and that all non-work time is devoted to the consumption of goods.[3] The 'full income' constraint is derived by combining the money and time constraints, (3) and (4) to give:

[3] An alternative formulation, yielding an equivalent empirical specification but a different conceptualisation of the observed lifestyle variables as fundamental commodities rather than as quantities of purchased goods, results from assuming that each lifestyle component is produced using goods and time inputs with a Leontief(fixed proportions) technology. In this case the observed levels of commodities are equivalent to the levels of the goods inputs. Assume that $C_j = f_j(G_j, t_j)$, for all commodities $j, j = 1...M$, where G_j and t_j are goods and time inputs to produce C_j units of commodity j. If each production function f_j is of the Leontief form where $C_j = min(G_j, \tau_j C_j)$, then the minimum cost of producing C_j is $(p_j + w\tau_j)C_j$. If each function f_j exhibits constant returns to scale, then the price of a unit of each commodity j is $(p_j + w\tau_j)$ and is independent of preferences. Note here that goods and time are not joint inputs for the commodities, j; inputs are commodity specific. Rather joint production occurs due to the production of health through the k lifestyle components. In this sense the goods G_j and τ_j are joint inputs in the production of the joint products $C_j, j = 1...k$ and health. Interestingly, in this case with a linear health production function the shadow price of health is also constant and equal to the minimum of the set $\{(1/\delta_j)(p_j + w\tau_j)\}$ where δ_j is the marginal product of C_j in $h(.)$. This case allows for the recovery of preference parameters when estimating a reduced form function for health and the commodities $C_j, j = 1...M$ using shadow prices. However, this approach to recovering preference parameters is not possible without information which enables the construction of shadow prices. Such information is not available in our dataset.

$$\sum_{j=1}^{M} (p_j + w\ \tau_j)C_j \leq m_0 + wT \ \equiv F \ , \tag{5}$$

where F is full income.

The shadow price of each good is dependent on the wage rate, and thus even with a common consumption technology (the set τ) across individuals, the shadow prices and relative shadow prices of lifestyle components will vary across individuals.

The Lagrangian function is:

$$\max_{C,H} \ \mathcal{L} = U(C,H) + \lambda \left((m_0 + wT) - \sum_{j=1}^{M} (p_j + w\tau_j)C_j \right) - \ \gamma(H - h(C)), \tag{6}$$

where λ and γ are the marginal utilities of full income and health respectively. The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial C_j} = \frac{\partial U}{\partial C_j} - \lambda(p_j + w\tau_j) + \gamma(\frac{\partial h}{\partial C_j}) = 0 \ \ \forall \ j = 1...M \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial H} = \frac{\partial U}{\partial H} - \gamma = 0 \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = m_0 + wT - \sum_{j=1}^{M} (p_j + w\tau_j)C_j \geq 0; \ \lambda \geq 0; \ \lambda \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = H - h(C) = 0 \tag{10}$$

Combining (7) and (8) for a particular C_j gives:

$$U_j + U_h h_j = \lambda(p_j + w\tau_j) \quad \forall j, j = 1...M \tag{11}$$

where U_j, U_h, and h_j denote partial derivatives.[4]

By solving the above equations we obtain the conditions for a solution in terms of the shadow price ratios and the total marginal utilities of goods i, j over all pairs of goods i, j. Ignoring corner solutions we obtain;

$$\frac{U_i + U_h h_i}{U_j + U_h h_j} = \frac{p_i + w\tau_i}{p_j + w\tau_j} \; \forall i, j \; where \; i \neq j \tag{12}$$

This is an obvious extension of the familiar and simple case involving no time prices and no joint production.

Solving the first order conditions (7) - (10) would give the Marshallian demands for health and the M goods in our system, the (unobservable) marginal utility of health (γ) at the optimum, and the (unobservable) marginal utility of full income (λ) at the optimum, in terms of the exogenous variables in the model:

$$C_i = f_i(Z) \; \forall i = 1...M, \tag{13}$$

$$H = h(Z) \; , \tag{14}$$

$$\lambda = \lambda(Z), \tag{15}$$

[4]In order to prevent indeterminacy in the solution of the model we assume that the budget constraint is always binding at an optimum; the marginal utility of full income is always positive. Otherwise, an individual will not wish to work to obtain more goods, and given that labour does not yield disutility, the choice variables in the model can take a range of values. For the budget to be exhausted at the optimum, we require that the left-hand side of (11) is positive for at least one good over the relevant domain of the consumption goods and health. For this to hold definitively we require some additional assumptions concerning each term of the left-hand side of (11). Specifically, we require that at least one good has a non-negative value of marginal utility and a positive marginal product of health over the whole domain, in addition to the assumption that the marginal utility of health is always positive. If the left-hand side of (11) is always negative for a given j, we would observe zero consumption of that good, as the right-hand side will always be greater than or equal to zero. This is a sufficient, although not a necessary condition for a binding income constraint at the optimum. So, the choice of C thus determines the total time spent in consumption, which determines L, and the amount of income earned wL, which together with unearned income and a binding budget constraint is just sufficient to buy the bundle chosen.

$$\gamma = \gamma(Z), \tag{16}$$

where Z consists of X, p, w, m_0 and τ. In general, the reduced forms do not separately identify preference and technological parameters. Furthermore, while they do provide information concerning the variables which influence lifestyle and health choices, they do not provide estimates of the impact of lifestyle choices on health. Equation (13) constitutes the foundations of our reduced form empirical models of lifestyle choices, with (13) used in estimation of the structural health production function, (2). The approach we implement here allows us to obtain estimates of the causal effects of lifestyle on health.

The transition from a theoretical to an empirical specification creates a number of problems. Firstly, we have concentrated on interior solutions when deriving the empirical model for a representative individual. However corner solutions are prevalent for most of the lifestyle components we consider. Secondly, in the theoretical model we have assumed that the measures of health and lifestyle are continuous variables. The dataset is not so accommodating. Thirdly, some of the variables which appear in the theoretical model are not measured in our sample and have to be proxied by other variables, or treated as inherently unobservable.

Our empirical model consists of a recursive structure with reduced form equations for the lifestyles(13) and the structural form of the health production function(2). Before outlining our estimation method and describing the results, we describe the dataset used to estimate the empirical model.

4 Data

The Health and Lifestyle Survey (HALS) is a national representative sample of adults living in private households in Great Britain. Carried out by Social and Community Planning research, the first wave of data (HALS1) were collected between Autumn 1984 and Summer 1985 during two home visits; firstly an hour long interview, followed by a nurse visit to collect physiological measurements and data on cognitive function. The available sample has information on 9003 individuals, although some gave incomplete responses. The response rate is 73.5% when considering the questionnaire alone. Additionally, as the HALS is a survey of private households, it may be particularly prone to selection bias in terms of assessing health and its interaction with behavioural indicators, as those individuals with severe or chronic health problems and disabilities are 'more likely to be in

hospital, or otherwise unavailable for interview'(Cox *et al.* (1987)). The survey sample has been compared to the 1981 Census of Population in order to gauge its representativeness. Among respondents who completed all three stages of the survey, there is a slight excess of women, particularly elderly women, and some under-representation of those households with lower incomes and education levels. However, the overall conclusion of the authors is that 'the study appears to offer a good and representative sample of the population.' The second wave of the data (HALS2) was collected in 1991 and includes 5352 interviewees. After deleting observations which have missing values at either the first or second wave of the survey for the variables in our model, we are left with complete observations on 4611 individuals.

4.1 Health and lifestyle variables

The endogenous behavioural variables employed are those which cover as many as possible of the lifestyle categories of the 'Alameda Seven' following findings from an epidemiological study of around 7000 individuals conducted in Alameda County, California in 1965. Subsequent analysis found that seven 'lifestyle' factors influenced physical health status (Belloc and Breslow (1972), Kenkel (1995)). These seven categories are; Diet, Smoking, Exercise, Alcohol, Sleep, Weight(for height), and Stress. Stress is excluded from our analyses due to the lack of a reasonable proxy in the dataset.[5] Weight(for height) is also excluded on the grounds that it does not have the status of a control variable, but rather is an intermediate health indicator.[6]

Health is measured by a binary indicator of self assessed health(SAH) relative to a representative individual of the respondent's own age(HEALTHEG respectively). The indicator takes the value one if an individual rates their health as excellent or good, and zero if they rate their health otherwise(fair or poor). While of interest in its own right, SAH is also rlated to number of other health measures. For example, Borg and Kristensen(2000) note that the vast majority of studies analysing the association between SAH and subsequent mortality find SAH to be a powerful predictor, even after conditioning on medical diagnoses and functional capacity. Burstrom and Fredlund(2001) find this predictive power to be stable across socio-economic groups. Self-reported health has also been used extensively in the health economics literature (e.g. Kemna(1987), Berger and

[5]The addition of social status indicators may attenuate the bias introduced by not considering stress as measures of social status are known to proxy stress levels (Wilkinson (1996),pp193-7).

[6]Falkner and Tanner (1986) claim that height is a particularly good indicator of past nutritional experience, while adult height is also recognised to be a predictor of morbidity and mortality risks (Waaler (1984), Steckel (1995)). Therefore we include height as a continuously measured exogenous variable in our econometric specifications.

Leigh(1989), Kenkel (1995)). Furthermore, a preliminary principal component analysis found that SAH correlates well with other indicators of health in the HALS.[7]

We use a binary variable (BREAKFAST) which equals one if an individual eats within one hour of waking and zero otherwise. A similar variable was employed by Kenkel (1995).

To measure smoking behaviour we also employ the binary variable (SMOKER) which equals one if an individual is a current smoker and zero otherwise.

The exercise variable is created for each individual by summing the time involved in each of fourteen exercise categories. Again we employ a binary variable(EXERCISE) which is equal to one if an individual participates in exercise and zero otherwise.

We measure alcohol consumption as a binary variable (ALQPRUD) which equals one if an individual drinks prudently and zero otherwise. This categorisation is gender specific and is based on the number of units consumed per week and medical advice at the time of HALS1. It takes the value one if a male drinks between 1 and 21 units of alcohol, and if a female drinks between 1 and 14 units of alcohol.[8]

Sleep was also recognised as a behavioural variable which affects health in the Alameda study. This is measured as an ordered categorical variable in HALS. However, Belloc and Breslow(1972) found that the healthiest number of hours of sleep was between seven and nine hours. Sleeping more or less than this 'optimum' reduced physical health. Therefore we created a binary variable to indicate either optimal or suboptimal sleeping levels. This variable(SLEEPGD) takes the value 1 if an individual sleeps the 'optimal' number of hours and takes the value 0 otherwise.[9]

4.2 Exogenous characteristics

We consider a general linear specification of the model to minimize the possibility of attributing variation in health to lifestyle factors, and thus to obtain estimates of causal

[7] A subjective measure may be of greater value in analysing the effects of health on lifestyle, while a more objective measure may be preferred if predicting service use for particular illnesses. Irrespective of the aims of a particular analysis however, a categorical measure is undoubtedly suboptimal. As Dasgupta (1993) notes, 'A person's state of health can take a continuum of values'.

[8] Reviewing epidemiological evidence Baum-Baicker(1985) concludes that moderate alcohol consumption can be beneficial to health, and thus we predict a non-linear relationship. This consideration helped determine our decision to exclude abstinence from the definition of prudent drinking. However, the measure that we use does not distinguish between a week of moderate drinking and one composed of periodic abstinence and heavier drinking. These two drinking styles are likely to have very different health effects. An alternative approach is the inclusion of a drinking style variable, either as a replacement for (ALQPRUD) or entering the empirical model as an additional endogenous variable. The former approach loses information on quantities, while the latter would necessitate the use of two alcohol consumption indicators. This complicates our methodology and introduces discrimination in the treatment of lifestyle components.

[9] We do not consider medical care as information is only available at HALS2.

effects.[10] The exogenous variables in the model can be grouped into categories which are considered in Table 5 along with definitions and sample statistics. As can be seen from the table we consider the following categories: social class, education, marital status, employment status, ethnic group, type of area, region, physical characteristics, tenure, household characteristics, and parental characteristics.[11]

4.3 Descriptive analysis

Table 6 presents HALS1 sample means for selected variables for sub-groups of lifestyle choices. In order to obtain a parsimonious and informative description of the samples we select sub-samples of the data based on the number of *a priori* 'healthy' behaviours.[12] Healthy behaviours are defined as sleeping well, taking breakfast, not smoking, consuming alcohol prudently, and exercising. The first column of sample statistics repeats the full sample means from Table 5. The second column gives sample means for those individuals whose observed behaviours are all unhealthy. The third column gives means for those who have one or two healthy behaviours, while the fourth column contains means for those who have three or four healthy behaviours. The final column contains sample means for those whose lifestyle can be considered as 'completely healthy' (i.e. 5 healthy behaviours). Columns 2-5 therefore define mutually exclusive and exhaustive sub-samples of the data. In the first row of Table 6 the sample sizes are shown for each sub-group along with the expected number of observations in each sub-groups if the behaviours were *randomly distributed*, (i.e. there is no tendency for healthy or unhealthy behaviours to be clustered either due to observed or unobserved characteristics). The expected values are obtained by calculating the relevant probabilities using the marginal means for the full sample and multiplying these values by the sample size.[13]

[10] The baseline individual is a member of the registrar general's social class classification (3) non-manual, has 'O' levels/CSE's or equivalent but no higher qualifications, is married, works full-time, is of white/european ethnic origin, lives in an inner city area in the south east of England, is female, lives in rented accomodation, and does not live with other smokers. Neither parent smoked.

[11] Wage rates are not measured, while income is reported categorically with substantial item non-response. We do not include income given the potential for sample selection bias and the reduction in sample size induced by using only those indivdiuals for whom we have income information. We attempted to impute missing values using a selectivity-corrected interval regression but were unable to obtain a satisfactory specification.

[12] To describe the data for each combination of lifestyle choices would require(potentially) $2^5 = 32$ sub-samples.

[13] More precisely, inverting the smoking variable so that it is equal to one when not smoking, and equal to zero when a smoker, a completely unhealthy individual will have a lifestyle outcome set equal to $\{0,0,0,0,0\}$. The probability of this set given the assumptions made in the text is $\prod_{j=1}^{5} Pr(Y_j = 0) = .01165$. Similarly, the probability of the set $\{1,1,1,1,1\}$ is $\prod_{j=1}^{5} Pr(Y_j = 1) = .06011$. The probability of one or two healthy behaviours can be written as $\sum_{j=1}^{5} Pr(Y_j = 1) \prod_{k \neq j} Pr(Y_k = 0) + \sum_{j=1}^{5} \sum_{k>j} Pr(Y_j =$

Examination of Table 6 suggests a number of observations. Firstly, the expected number of individuals behaving completely healthily or unhealthily are around half of what we actually observe. This suggests that health-related behaviours are not randomly distributed, but rather that healthy and unhealthy behaviours cluster together in certain individuals. This clustering may be due to observed or unobserved factors. Secondly, average SAH gradually increases as the number of healthy behaviours increases.

We can examine how the clustering of behaviours is related to exogenous characteristics by examining how average characteristics vary across sub-samples. For example, it can be seen that the proportion of individuals in higher social groups(regsc1s and regsc2) gradually increases as we move from completely unhealthy to completely healthy lifestyles. Conversely, the proportions in lower social class groups gradually decrease. For educational attainment a similar pattern exists: as we move from a completely unhealthy to a completely healthy lifestyle the proportions of individuals with higher levels of education(lhqa, lhqhnd, lhqdg) gradually increases while the proportions of individuals with lower levels of education(lhqnone and baseline) gradually decrease.

For the employment status variables, we observe that part-time workers are more likely to have a healthy lifestyle , while those that are sick are more likely to have an unhealthy lifestyle. Those that are retired are more likely to have an intermediate lifestyle rather than a completely healthy or unhealthy one, while students are more likely to have a healthy than an unhealthy lifestyle, while those that keep house are more likely to have an unhealthy one. The same is true, and to a stronger degree for those that are unemployed and who work shifts.

There is no discernible trend for gender or height while those with a completely healthy lifestyle are more likely to be younger than average. Houseowners are more likely to be unhealthy than others as are those who live with other smokers. Similarly, those individuals whose parents both smoked are more likely to have an unhealthy lifestyle, although there is some evidence of a different behaviour pattern depending on whether an individual's mother or father smoked.

Some comments are in order concerning the above observations. Firstly, while for some variables there is a strong gradient as we move from an unhealthy to a healthy lifestyle, this does not indicate a *causal effect* of that exogenous variable: these are simple correlations only. Secondly, these observations do not relate directly to specific health-related behaviours, but are for the average lifestyle within each category. Therefore, there is no reason to expect these gradients to exist for each health-related behaviour. However, these

1)$Pr(Y_k = 1)\prod_{l\neq k\neq j} Pr(Y_l = 0) = .09751 + .25014 = .34765$. Finally, due to mutual exclusivity and exhaustivity, the probability of 3 or 4 healthy behaviours is equal to $1 - Pr(0) - Pr(1/2) - Pr(5) = .58060$.

results are suggestive of effects we may observe in reduced form models for the health-related behaviours. Thirdly, a further complication for interpretation of the observed health gradient is the expected endogeneity of the health-related behaviours in the health production function. Thus, while it may hold that partial correlations between each lifestyle indicator and exogenous characteristics confirm the observations above, structural estimation of the SAH model need not exhibit the effect of healthy or unhealthy lifestyles observed above. There are two reasons why this may occur. Firstly, variation in SAH across lifestyle categories which appears in Table 6 may be explained by exogenous characteristics: lifestyle has no independent effect on SAH. Secondly, although partial correlations between SAH and health-related behaviours may indicate the existence of effects of health-related behaviours conditional on exogenous characteristics, these correlations may be due to unobserved factors which affect SAH and which are correlated with health-related behaviours. In this case, an estimation method which adequately accounts for these unobservable factors may lead to estimates of structural effects of health-related behaviours which are statistically and quantitatively insignificant.

5 Estimation Strategy

A consistent estimator of the health production function must account for the endogeneity in equation (2), introduced by the existence of lifestyle components as regressors and correlations between the errors of the models determining lifestyle choices and that which determines self-assessed health. In the absence of the complication that our endogenous variables are binary, many easy to implement estimators are available for the linear model such as 2SLS, 3SLS, and Full Information Maximum Likelihood(FIML).

However, in the nonlinear case, the efficient ML estimator in a model with more than 3 endogenous binary variables with correlated error terms has until recently been computationally intractable. However, advances in simulation-based estimation and inference allow for consistent estimators which are, under certain conditions, asymptotically equivalent to ML(see e.g. Hajivassiliou and Ruud(1994), Gourieroux and Monfort(1996)). Here we use Maximum Simulated Likelihood(MSL) to obtain consistent and asymptotically efficient estimators of the parameters of both the reduced form system of health-related behaviours and the structural model simultaneously. This is achieved by assuming that the error terms of the full recursive system consisting of the reduced form equations for the health-related behaviours, (13) and the structural form of the health production function, (2), have a multivariate normal distribution. This allows us to estimate the model using a multivariate probit formulation where the correlations between the error terms are

unrestricted.[14]

In particular, we consider the structural health production function for SAH at HALS2 to be a function of HALS1 lifestyles and HALS2 exogenous variables, with HALS1 lifestyles a function of HALS1 exogenous variables. As no endogenous variable is observed at more than one time point in this formulation, the model can be estimated as a cross-sectional multivariate probit model with zero restrictions on the parameters of the equations determining outcomes. The correlation parameters indicate whether and how unobservable factors jointly affect lifestyle decisions and health outcomes.[15]

Implementation of MSL estimation of a multivariate probit model with many outcomes and a complex error structure requires a simulator for the probabilities that enter the log-likelihood function. There are many alternatives available for the simulation of multivariate normal rectangle probabilities(e.g. Hajivassiliou et al.(1996) and Vijverberg(1997)). Given the current Monte-Carlo evidence (e.g. Hajivassiliou et al.(1996) and Vijverberg(1997)), and the theoretical properties derived by Borsch-Supan and Hajivassiliou (1993), the leading simulator is the Smooth Recursive Simulator (SRC), or Geweke-Hajivassiliou-Keane (GHK) simulator. This simulator is strictly bounded by zero and one, smooth in the parameters (conditional on a smooth method to obtain the inverse of the truncated normal CDF), unbiased, and consistent in the number of replications R. Monte-Carlo evidence also shows that it has low variance.[16]

To further reduce the variance of the simulators we use antithetic acceleration. These simulators use the original set of uniform random draws along with their reflections to

[14] The correlation matrix is identifiable in the multivariate probit model, while the covariance matrix is not(see e.g Chib and Greenberg(1998)). This situation parallels parameter identification in the univariate probit model where only scaled versions of the β's are identified. In the multivariate probit model the correlation matrix and the β's are identifiable, but not the β's and the covariance matrix.

[15] A simultaneous equation system for a set of binary outcome variables requires certain restrictions for coherency and identification. The condition for coherency is described by Schmidt(1981), and requires the structure of the model to be recursive. Our system satisfies this condition(Schmidt Condition 12.5) as a special case of a simultaneous equation system with parametric restrictions. In this recursive structure certain restrictions are also required for identification. In particular, a model without exclusion restrictions in the structural equation of interest and a correlated error structure is not identified.(Maddala and Lee(1976), Maddala(1983)p122-3). Identification is provided here by exclusion restrictions and by variation over time in the exogenous characteristics which affect lifestyle choices and also directly affect health.

[16] Consistency can be obtained by reducing the error of the simulated sample log-likelihood to zero, as the number of replications, R, $\rightarrow \infty$, at a sufficient rate with the sample size, N. For a finite variance and unbiased probability simulator, as the number of replications grows the bias and variance of the approximation to the sample log-likelihood approach zero, but the variance must reduce at a sufficient rate to avoid asymptotic bias in the limiting distribution of the estimator of the parameters of the model, $\widehat{\theta_{MSL}}$. Hajvassiliou and Ruud(1994) show that a sufficient rate for this is $R/\sqrt{N} \rightarrow \infty$ as $N \rightarrow \infty$. Furthermore, they show that this rate is also sufficient for MSL to be asymptotically efficient such that no correction is required for the covariance matrix relative to that obtained for maximum likelihood: 'Given enough simulations to overcome bias, there are enough simulations to make the asymptotic contribution of simulation to the limiting distribution of $\widehat{\beta_{MSL}}$ negligible.' (Hajivassiliou and Ruud(1994))

estimate the probability of the observed outcomes for each individual. Here $\tilde{l}(\theta; y_i, X_i, \xi_i)$ is the simulated likelihood contribution of individual i at parameters θ given outcomes y_i and data on exogenous characteristics X_i. ξ_i are draws from the standard uniform distribution. The simulator is given by:

$$\tilde{l}(\theta; y_i, X_i, \xi_i) = (1/2R) \sum_{r=1}^{2R} \tilde{l}(\theta; y_i, X_i, \xi_{ir}) \tag{17}$$

$\xi_i = \xi_{i1}, \ldots, \xi_{i2R}$ and $\xi_j = -\xi_{j-R}$ for $j = R+1, \ldots, 2R$. [17]

In addition to estimation of the full recursive model, we also consider two simpler estimators which allow us to evaluate the impact of endogeneity of lifestyle on the parameter estimates. These are obtained by estimating a univariate probit model for the health outcome based on the same specification as was used for the health equation in the recursive system. However, here we do not allow for correlation between the errors of the health equation and the lifestyle indicators and hence do not control for endogeneity. Thus comparing estimates of the parameters of the health equation in the recursive system and in the univariate model allows us to estimate the effects of not controlling for endogeneity. Also, a test for exogeneity is provided by estimation of the recursive system: Wald tests that the correlations between the errors of the health equation and the lifestyle equations are equal to zero are immediate tests of exogeneity.

The second estimator we consider is for a model of SAH at HALS2 which excludes the lifestyle indicators. This is estimated as a univariate probit. Comparison of these results with those from the recursive system allow us to consider the effect of lifestyle factors on the impact of other variables on health, such as socio-economic group. These results are of interest as they allow us to make statements concerning the degree to which health inequalities across socio-economic groups are due to lifestyle choices.

6 Results

6.1 Lifestyle equations

Table 1 shows selected marginal effects for the reduced form lifestyle models estimated using the MVP specification of the full recursive system.[18] A slight gradient in the prob-

[17] Hajivassiliou(2000) presents Monte-Carlo evidence suggesting that the antithetically accelerated simulator for multivariate normal rectangle probabilities is superior to the standard SRC simulator.

[18] Full results for the coefficients of the lifestyle equations are presented in Table 7. While the effects are average effects for the dummy variables we refer to all effects as marginal. All marginal effects throughout the discussion of the results were calculated at the means of the other variables in the model.

ability of sleeping well can be observed by social class, with those in the highest socioeconomic group significantly more likely to sleep well than those in the baseline category (skilled manual).[19] Those from lower social classes are less likely to eat breakfast.[20] We find a significant and strong social class gradient for the probability of being a smoker. While this association has been observed in other studies(e.g Borg and Kristensen(2000), Lynch *et al.*(1997)), most have controlled for only a few other variables in measuring these associations. We find that a significant relation with social class remains after controlling for a large number of variables. Similar observations can be made for prudent alcohol consumption and for exercise. In both cases, it is the higher social classes who are more likely to behave 'healthily'.

We noted in the descriptive analysis that healthy behaviours tend to cluster, and that individuals who behave 'healthily' are more likely to be found in higher social classes. This observation remains after accounting for other variables: those in lower social classes are, conditional on other observed variables, more likely to behave unhealthily in all measured respects relative to those from high social classes.[21] This accords with the perspective of Evans et al.(1994) who argue that health related behaviours should be seen as responses to environmental conditions. There is no implication here that preferences are irrelevant, but rather that they too are conditioned by the social environment of an individual.

Education does not significantly and independently affect the probability of sleeping well, while those with no qualifications are significantly less likely to eat breakfast and those with a degree are significantly more likely to do so. The more educated are less likely to smoke, while those with no qualifications are significantly less likely to be prudent alcohol consumers and to exercise than those with school-leaving qualifications. While a gradient for educational qualifications is less visible than for social class, and the marginal effects are not as large, educational differences appear to relate, conditional on other variables, to the clustering of healthy and unhealthy behaviours by educational status that we observed in the descriptive analysis. In general, significant results for the other variables accord

[19] Throughout the discussion of results we use 'significant' and 'significantly' with reference to statistical significance, not practically important or large in magnitude.

[20] Of recent studies that have considered more than one health-related behaviour, Lynch *et al.*(1997) used Finnish data from the Kuopio Ischaemic Heart Disease Risk factor study and found that, conditional on age, white collar workers had the lowest mean consumption of cigarettes for current smokers.

[21] We should interpret these estimates carefully. Other potentially unobserved variables such as childhood circumstances may lead to indirect selection(due to unobserved heterogeneity) such that individuals with 'deprived' backgrounds may be both more likely to behave poorly and be in lower social classes. This selection can take the form of preference and constraint variation. While we do not take into account the potential endogeneity of social class, we do condition on education and a number of parental characteristics. Similar considerations apply to the education and work status variables. Also of concern is the possibility of direct selection(reverse causality) where lifestyle variables directly affect social class or other variables which are considered exogenous.

with the simple correlations observed in the descriptive analysis.[22]

Table 2 shows the correlation matrix for the full recursive model. Ignoring the health column of the table for now, we concentrate on the significant correlations in the remainder of the matrix. Unobservables which affect the propensity to eat breakfast are positively related to those which affect sleeping well, not smoking, being a prudent alcohol consumer and exercising. This observation further helps to explain the observed clustering of health-related behaviours. From these correlations it appears that the unobserved propensity for a 'healthy' lifestyle is predicted well by unobserved characteristics which determine whether an individual eats breakfast. In conjunction with the positive correlations for prudent alcohol consumption and exercise and the negative correlations between smoking and exercise, we can interpret this matrix as indicating that individuals have unobserved propensities for leading completely 'healthy' or 'unhealthy' lifestyles. While, by definition, these propensities are not observed, the clustering of behaviours observed in the descriptive analysis appears to be partially due to correlated unobservables which affect the relative utilities of behaving healthily and unhealthily for each lifestyle choice. These may in turn be related to differences in childhood circumstances, attitudes to risk and the rate of time preference. Barsky *et al.*(1997) offer some evidence for the impact of risk attitudes on lifestyle choices using experimental data. They find that risk tolerance is positively related to risky behaviours such that risk tolerance is a statistically significant and quantitatively important factor in explaining whether an individual is a heavy drinker or a current smoker, even after conditioning on demographic characteristics. They also cite studies which report a biological basis for characteristics which appear to be related to risk tolerance. These suggest causal effects and that general personality characteristics affect different choices in a similar way. Barsky *et al.*(1997) also find that the wealthy and highly educated are more risk tolerant than average suggesting that the socio-economic gradients in behaviour found here and elsewhere are not explained by different risk attitudes.

Unobservable heterogeneity may also reflect correlations in the (perceived) marginal products of lifestyles with respect to health, due, for example, to differences in health knowledge(Kenkel(1991)). They may also reflect different opportunity costs, in terms of forgone income, due to the unobserved wage rate and time costs of each lifestyle choice.

[22] Interpretation of the effect of the sick variable as causal is particularly inappropriate. However, to prevent the effect of the variable being subsumed in the intercept it was necessary to include it explicitly as one of the work status variables.

Table 1: Selected marginal effects for reduced form lifestyle models

	Sleepgd	Breakfast	Smoker	Alqprud	Exercise
Social Class					
regsc1s	.09*	.04	-.06	.02	.07*
regsc2	.03	-.04	-.01	-.02	-.01
regsc3a	-.02	-.07*	.07*	-.06*	-.04
regsc4	-.01	-.09*	.07*	-.08*	-.09*
regsc5n	-.01	-.06	.09*	-.17*	-.10*
Education					
lhgdg	-.05	.07*	-.07*	.03	.03
lhqhnd	-.02	.06*	-.04	.004	-.02
lhqA	-.02	.03	.01	-.02	-.02
lhqnone	-.02	-.11*	.05*	-.06*	-.10*
lhqoth	-.001	-.07	.06	-.06	.02
Work					
part	.08*	.11*	-.03	.0004	.02
sick	-.16*	-.11	.16*	-.31*	-.16*
retd	.06	.17*	.05	-.03	.02
stdnt	.10	.11	-.08	-.004	.02
keephse	-.004	.04	.003	-.09*	-.07*
unemp	-.004	-.02	.15*	-.08*	-.06
wkshft	-.13*	-.07*	.08*	-.04	.02

Table 2: Correlations for full recursive model estimated by MSL(R=50)

	HEALTHEG	SLEEPGD	BREAKFAST	SMOKER	ALQPRUD	EXERCISE
HEALTHEG	1.00					
SLEEPGD	.06	1.00				
BREAKFAST	-.18	.12*	1.00			
SMOKER	.21*	-.04	-.29*	1.00		
ALQPRUD	-.75*	.03	.11*	-.04	1.00	
EXERCISE	.07	.03	.09*	-.09*	.14*	1.00

Table 3: Selected marginal effects for alternative models of SAH

	MVP(R=50)	Exogenous	Excluded
Social Class			
regsc1s	.07*	.09*	.09*
regsc2	.02	.03	.03
regsc3a	.02	-.01	-.02
regsc4	.01	-.04	-.05*
regsc5n	.003	-.06	-.08*
Education			
lhgdg	-.03	-.01	-.003
lhqhnd	-.01	-.003	-.003
lhqA	.001	-.001	.001
lhqnone	-.06*	-.10*	-.11*
lhqoth	-.01	-.02	-.02

6.2 SAH Equations

We begin by considering the results of the model for SAH at HALS2 where lifestyle factors measured at HALS1 are excluded.[23] The 'excluded' column of Table 3 presents marginal effects for the social class variables using a simple univariate probit model without the lifestyle variables. It is immediately apparent that a significant social class gradient in SAH exists conditional on education, marital status, ethnic status, type of area, gender, height, and a flexible function of age. For each social class, the probability of reporting excellent or good SAH is greater than for the class immediately below it. These differentials are such that those in the highest social class are significantly more likely to report excellent or good SAH relative to the baseline category, while those in the two lowest social classes are significantly more likely than the baseline category to report fair or worse health. This gradient is emphasized by the graphical illustration in Figure 1. A gradient is not clear for education, with only those with a degree or higher degree having a significantly different probability of reporting excellent or good health relative to those in the baseline category. While these results are indicative of a direct impact of socio-economic characteristics on health, this model omits the lifestyle variables. As reported in the previous section, there are significant effects of socio-economic status and other variables on lifestyle choices at HALS1. By including the lifestyle variables we are able to obtain estimates of the effects of HALS1 lifestyle choices on HALS2 SAH and examine how including lifestyle affects the estimates of the impact of social class and education.

The 'exogenous' column of Table 3 contains marginal effects for the social class and education dummies on the probability of reporting excellent or good SAH when including the HALS1 lifestyle variables(but assuming that they are exogenous) using a univariate probit model. A social class gradient appears to exist, but the effects are not as strong for the lowest social classes relative to the baseline category and are now insignificant. The

[23] Full results for all specifications of the SAH models are contained in Table 8.

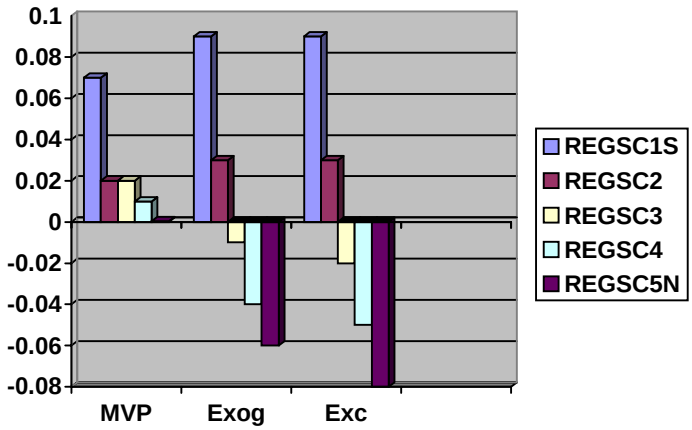

Figure 1: Social Class Gradients for alternative models of SAH

effects of education are almost identical to those obtained when excluding the lifestyle variables. Epidemiological studies have also examined the impact of lifestyle on the social class gradient in SAH and other health indicators, including mortality and specific disease indicators such as heart disease (Borg and Kristensen(2000), Power *et al.*(1998), Marmot *et al.*(1996), Lynch *et al.*(1996)). Using Danish data from a random sample of individuals interviewed in 1990 and 1995, Borg and Kristensen used simple logit models to estimate the odds ratios for reporting a reduction in SAH in 1995 conditional on having good or very good SAH in 1990. They found that 17% of the odds ratio for the highest status occupational group relative to the lowest was explained by tobacco consumption and an indicator for obesity, after controlling for age, gender and an indicator of illness. However, tobacco consumption led to only a 4% reduction.[24] Lynch *et al.*(1996) using the Kuopio Ischaemic Heart Disease Risk Factor Study data, found that 35% of the relative risk of all-cause mortality for the lowest income quintile relative to the highest could be attributed (conditioning only on age) to behavioural factors (smoking, alcohol consumption, and physical activity).

We use a similar measure of the effect of lifestyle on the social class gradient in health

[24] Borg and Kristensen and the other studies cited used the same measure of the effect of including lifestyle and other variables on the social class gradient. Defining OR_E as the odds ratio for the lowest social class relative to the highest without lifestyle and other variables and OR_I as the odds ratio for the lowest social class relative to the highest including lifestyle and other variables, the measure used is $\frac{OR_I - OR_E}{OR_E - 1}$. This is the percentage reduction in the odds ratio due to measured lifestyle factors.

Table 4: Lifestyle marginal effects for alternative models of SAH

	MVP(R=50)	Exogenous
Lifestyle		
Sleepgd	-.003	.05*
Breakfast	.05	-.01
Smoker	-.18*	-.11*
Alqprud	.38*	.03*
Exercise	-.05	.04*

by computing the percentage change in the difference between the marginal effects for the highest and lowest social classes and education groups. This measure is the percentage change in the difference in the probabilities of reporting an excellent or good SAH comparing the highest and lowest social classes.[25]

Comparing the case where the lifestyle variables are excluded with that where lifestyle variables are included but assumed exogenous leads to a reduction in the social class gradient of 12%. For the education variables, the figure is 18%. This is found by comparing those with no qualifications to those with a degree or higher degree.[26] These results are comparable to those found in the epidemiological literature where it has been concluded that lifestyle factors do not account for much of the social class gradient in health, particularly when used in a multivariate analysis with other potential mediators of the relationship between socio-economic status and health. Power *et al.*(1998) used the 1958 British birth cohort and estimated the contributions of various factors including childhood and adolescent variables to the social class gradient using the odds ratio approach described above and a series of logit models. Allowing for smoking status at ages 23 and 33 in a model for SAH at age 33 changed the odds ratio by 25%, while alcohol, diet and body-mass index had small effects. These effects were found in the absence of any other explanatory variables, although adult smoking continued to have an effect, 14%, conditional on a number of variables recording behaviour and other characteristics earlier in life. Power *et al.* suggest a number of reasons for the small effects of the other variables, particularly the negligible effect for a dummy variable for alcohol consumption similar to ours. Two proposed causes were the lack of association of lifestyle factors with social class or because the age of the sample was insufficient for health effects to be manifest. Neither of these explanations is reasonable here given the results in Tables 1 and 4.

Table 4 shows marginal effects for the impact of the lifestyle variables measured at

[25] Other measures are of course possible, such as the variance of the effects or the average pairwise difference comparing each group with that immediately below it, but our measure is closer to those used in the epidemiological literature, thus providing us with better comparators.

[26] Education is often perceived as a good indicator of permanent income, while social class indicates occupational status and the work environment.

HALS1 on SAH at HALS2. The exogenous column suggests that sleeping a 'healthy' number of hours, prudent alcohol consumption and exercise have relatively small but significant positive effects on the probability of reporting excellent or good SAH, while smoking has a significant and large negative marginal effect on the probability of reporting excellent or good SAH. Taking breakfast is essentially irrelevant for SAH, as found by Kenkel(1995). Thus, it appears that while social class at HALS1 and lifestyle are significantly related, and lifestyle at HALS1 and SAH at HALS2 are also significantly related, the observed relationship between social class at HALS2 and SAH is not due to these correlations.

However, examination of the MVP column of Table 4 shows that accounting for unobserved heterogeneity changes the estimates substantially. Sleeping 'healthily' is now estimated to be largely irrelevant for SAH while exercise is now estimated to have a negative but insignificant impact. Taking breakfast remains unimportant. Kenkel(1995) estimated 'healthy' sleeping and exercise to be beneficial assuming these characteristics were exogenous, while also finding breakfast to be unimportant. Our results demonstrate the importance of allowing for unobservable heterogeneity.[27] The most dramatic effects occur in the absolute values of the marginal effects for smoking and alcohol consumption. Smoking is now estimated to be more damaging for SAH than when considered exogenous, as is imprudent alcohol consumption. For smoking the marginal effect changes by 64 %, while the effect of prudent drinking is more than 10 times the magnitude estimated when it is considered exogenous. These changes can be explained by considering the HEALTHEG column of Table 2. While sleep, breakfast, and exercise are estimated to be exogenous in the SAH equation at the 5 % level, the null of exogeneity for the smoking and alcohol variables is rejected. These correlations drive the changes in the results we observe. In particular, those with unobserved characteristics which lead them to smoke are more likely to have unobserved characteristics which induce a high level of SAH, while individuals with unobserved characteristics which increase the probability of prudent drinking are more likely to have unobserved characteristics which decrease the probability of reporting excellent or good SAH. These observations suggest that the effects we observe for the univariate probit model are underestimates of the effects which would be estimated by randomized treatment assignment to smoking or alcohol consumption.[28] Two explanations for this result are apparent which we term *ex ante* and *ex post* selection.[29]

[27] It should be noted that the coefficient standard errors in the MVP model are between 2 and 5 times those in the univariate probit model. This is expected as identification is provided by pure exogenous variation in the lifestyle variables.

[28] Changes in the effects of the other variables are driven by indirect effects, such as correlations with smoking and alcohol consumption.

[29] Alternative explanations are offered by Fuchs(1982) who offers some evidence of positive and negative

The *ex ante* explanation is that those individuals who choose to smoke and drink imprudently recognize that their latent health status is high and prefer to trade off health for consumption of these health depreciating goods. It is important to recognize our conception of latent health status is observable by neither the researcher nor the individual, but is a noisy estimate of future health status based on information obtained by the individual, such as parental and grandparental ages and causes of death. Some evidence for this explanation is provided by Hurd and McGarry(1995) who find that individual estimates of survival probabilities aggregate closely to life table values and covary appropriately with known risk factors. They also find that individuals are able to appropriately update subjective survival probabilities based on new information such as the death of a parent, and that the subjective survival probabilities predict actual survival. The *ex post* explanation instead relies on the observation that individuals are able to observe their current health status and update estimates of the marginal benefits of behaving healthily at each decision point.[30]

The role of unobservable heterogeneity offers an explanation for the small impact of lifestyle on the social class and educational status gradients found in the univariate models and the epidemiological literature. Table 3 and Figure 1 show that accounting for the endogeneity of lifestyle factors dramatically reduces the social class gradient. As for the exogenous case, the MVP estimates indicate that only those in the highest social class have a significantly different probability of reporting excellent or good SAH relative to the baseline category, but now the gradient across the other social classes is much smaller. While comparing the model where lifestyle was considered exogenous with that where lifestyles were excluded suggested a 12 % reduction in the social class gradient due to lifestyle, comparing the excluded and MVP columns gives the total reduction in the gradient as 59%. 47% of the effect of lifestyle on the social class gradient is due to the effects of unobserved heterogeneity. This effect is hidden in the exogenous model. A similar decomposition for education finds that the total reduction in the educational gradient is 73%, with 55% of the effect of lifestyle on the education gradient due to the effects of unobserved heterogeneity.

relationships between the rate of time preference and smoking decisions and health respectively, and Kenkel(1991) who proposes that health knowledge may explain the differential consumption patterns of those with different education levels.

[30] While we have a measure of SAH at HALS1, we do not estimate the MVP model including this variable in the HALS1 lifestyle equations due to simultaneity problems and the lack of individual information prior to HALS1.

7 Conclusion

We developed a simple model to identify interactions between health related behaviour and self-assessed health status, given other observable and unobservable factors. Unobservable heterogeneity may reflect underlying causal factors such as correlations in the direct marginal utilities of health, income, and lifestyle choices which may in turn be related to differences in genetic characteristics, childhood circumstances, attitudes to risk and the rate of time preference. They may also reflect correlations in the marginal products of lifestyles with respect to health.

We estimated the structural parameters of a health production function, together with the reduced form parameters for the lifestyle equations using panel data from the *Health and Lifestyle Survey*(HALS) conducted in the United Kingdom in 1984 and 1991. This is achieved using Maximum Simulated Likelihood for a multivariate probit(MVP) model with discrete indicators of lifestyle choices and self-assessed health(SAH).

We find that prudent drinking and not smoking in 1984 have dramatic positive effects on the probability of reporting excellent or good SAH in 1991. The failure of epidemiological analyses to account for unobserved heterogeneity can explain their low estimates of the relevance of lifestyle in the relationship between socio-economic status and health. While comparing the model where lifestyle was considered exogenous with that where lifestyles were excluded suggested a 12 % reduction in the social class gradient due to lifestyle, comparing the model with lifestyles excluded and the MVP model gives the total reduction in the gradient as 59%. 47% of the effect of lifestyle on the social class gradient is due to the effects of unobserved heterogeneity. This effect is hidden in the exogenous model. A similar decomposition for education finds that the total reduction in the educational gradient is 73%, with 55% of the effect of lifestyle on the education gradient due to the effects of unobserved heterogeneity. Accounting for unobserved heterogeneity also leads us to conclude that indicators for sleep, exercise, and breakfast in 1984 are unimportant for SAH in 1991.

References

Auster, R., I.Leveson, and Sarachek, D. (1969). The production of health: An exploratory study. *Journal of Human Resources*, 4:412–436.

Barsky, R., Juster, F., Kimball, M., and Shapiro, M. (1997). Preference parameters and behavioural heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics*, 112(2):537–579.

Becker, G. (1965). A theory of the allocation of time. *Economic Journal*, 75:493–517.

Belloc, N. and Breslow, L. (1972). Relationship of physical health status and health practices. *Preventive Medicine*, 1:409–421.

Berger, M. and Leigh, J. (1989). Schooling, self-selection, and health. *Journal of Human Resources*, 24:433–455.

Birch, S. and Stoddart, G. (1989). Incentives to be healthy: an economic model of health-related behaviour. Technical report, Centre for Health Economics and Policy Analysis working paper series 24, Ontario, Canada.

Blaylock, J. and Blisard, W. (1992). Self-evaluated health status and smoking behaviour. *Applied Economics*, 24:429–35.

Borg, V. and Kristensen, T. (2000). Social class and self-rated health:can the gradient be explained by differences in life style or work environment? *Social Science and Medicine*, 51:1019–1030.

Borsch-Supan, A. and Hajivassiliou, V. (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics*, 58:347–368.

Burstrom, B. and P.Fredlund (2001). Self-rated health:is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Epidemiology and Community Health*, 55:836–840.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85(2):347–361.

Cox, D., Blaxter, M., Fenner, J., Golding, J., and Gore, M. (1987). *The Health and Lifestyle Survey*. Health Promotion Research Trust, London, first edition.

Dardanoni, V. and Wagstaff, A. (1987). Uncertainty and the demand for health care. *Journal of Health Economics*, 9:23–38.

Dasgupta, P. (1993). *An enquiry into well-being and destitution*. Oxford University Press, New York, first edition.

Ehrlich, I. and Chuma, H. (1990). A model of the demand for longevity and the value of life extension. *Journal of Political Economy*, 98:761–782.

Evans, R., Barer, M., and Marmor, T. (1994). *Why are Some People Healthy and Others Not? The Determinants of Health of Populations*. Walter de Gruyter, New York, first edition.

Falkner, F. and Tanner(eds), J. (1986). *Human Growth: A Comprehensive Treatise, volume 3*. Plenum Press, New York, second edition.

Forster, M. (2001). The meaning of death: some simulations of a model of healthy and unhealthy consumption. *Journal of Health Economics*, 20:613–638.

Fuchs, V. (1982). *Time preference and health: an exploratory study*, pages 93–120. University of Chicago Press. in Fuchs, V. Economic aspects of health.

Fuchs, V. (1986). *The Health Economy*. Harvard UP, Cambridge, Massachusetts, first edition.

Gourieroux, C. and Monfort, A. (1996). *Simulation-based econometric methods*. Oxford University Press.

Gravelle, H. and Rees, R. (1992). *Microeconomics*. Longman, London, second edition.

Grossman, M. (1971). The economics of joint production in the household. NBER paper 7145, Center for Mathematical studies in Business and Economics, University of Chicago.

Grossman, M. (1972). The demand for health: a theoretical and empirical investigation. Technical report, National Bureau of Economic Research Occasional Paper 119, New York.

Hajivassiliou, V. (2000). *Some practical issues in maximum simulated likelihood*, chapter 3, pages 71–99. CUP. in Simulation-based inference in econometrics:Methods and Applications, edited by R Mariano, T Schuermann, and MJ Weeks.

Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics*, 72:85–134.

Hajivassiliou, V. and Ruud, P. (1994). *Classical estimation methods using simulation*, pages 2383–2441. Amsterdam:North Holland. in 'Handbook of Econometrics Vol 4' edited by R. Engle and D. McFadden.

Hu, T.-W., Ren, Q.-F., Keeler, T., and Bartlett, J. (1995). The demand for cigarettes in California and behavioural risk factors. *Health Economics*, 4:7–14.

Hurd, M. and McGarry, K. (1995). Evaluation of subjective probability distributions in the health and retirement study. *Journal of Human Resources*, 30(S):S268–S292.

Kemna, H. (1987). Working conditions and the relationship between schooling and health. *Journal of Health Economics*, 6:189–210.

Kenkel, D. (1991). Health behaviour, health knowledge and schooling. *Journal of Political Economy*, 99:287–305.

Kenkel, D. (1995). Should you eat breakfast? Estimates from health production functions. *Health Economics*, 4:15–29.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74:132–157.

Lantz, P., House, J., Lepkowski, J., Williams, D., Mero, R., and J.Chen (1998). Socioeconomic factors, health behaviours and mortality. *Journal of the American Medical Association*, 279(21):1703–1708.

Lynch, J., Kaplan, G., Cohen, R., Tuomilehto, J., and J.T.Salonen (1996). Do cardiovascular risk fctors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? *American Journal of Epidemiology*, 144:934–942.

Lynch, J., Kaplan, G., and J.T.Salonen (1997). Why do poor people behave poorly? variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Social Science and Medicine*, 44:809–819.

Maddala, G. (1983). *Limited Dependent and Qualitiative variables in econometrics.* Cambridge University Press, Cambridge, 1st edition.

Maddala, G. and Lee, L.-F. (1976). Recursive models with qualitative endogenous variables. *Annals of Economic and Social Measurement*, 5(4):525–545.

Marmot, M., Ryff, C., Bumpass, L., Shipley, M., and Marks, N. (1997). Social inequalities in health:next questions and converging evidence. *Social Science and Medicine*, 44:901–910.

McGinnis, J. and W.H.Foege (1993). Actual causes of death in the United States. *Journal of the American Medical Association*, 270:2207–2212.

Mullahy, J. and Portney, P. R. (1990). Air pollution, cigarette smoking, and the production of respiratory health. *Journal of Health Economics*, 9:193–205.

Pollak, R. and Wachter, M. (1975). The relevance of the household production function and its implications for the allocation of time. *Journal of Political Economy*, 83:255–277.

Power, C., Matthews, S., and O.Manor (1998). Inequalities in self-rated health: explanations from different stages of life. *The Lancet*, 351:1009–1014.

Rosenzweig, M. and Schultz, T. (1983). Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of Political Economy*, 91:723–746.

Schmidt, P. (1981). *Constraints on the parameters in simultaneous Tobit and Probit models*, chapter 12, pages 422–434. MIT press: Cambridge,MA. in Stuctural analsysis of discrete data with econometric applications, edited by C.F. Manski and D. McFadden.

Steckel, H. (1995). Stature and the standard of living. *Journal of Economic Literature*, 33:1903–1940.

Suen, W. and Mo, P. (1994). Simple analytics of productive consumption. *Journal of Political Economy*, 102:372–383.

Vijverberg, W. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics*, 76:281–307.

Waaler, H. (1984). Height, weight and morbidity: The Norwegian experience. *Acta Medica Scandanavia (supplement)*, 679.

WHO, H. E. U. (1986). Lifestyles and Health. *Social Science and Medicine*, 22:117–124.

Wilkinson, R. (1996). *Unhealthy Societies. The Afflictions of Inequality*. Routledge, London and New York, first edition.

Table 5: Variable Definitions and Sample Statistics

Variable	Definition	Mean(HALS1)	SD(HALS1)	Mean(HALS2)	SD(HALS2)
Health and Lifestyle					
Health	1 if SAH excellent or good, 0 otherwise	.754	.430	-	-
Sleepgd	1 if sleep between 7 & 9 hrs,0 otherwise	.624	.485	-	-
Breakfast	1 if eats within 1 hr of waking,0 otherwise	.655	.475	-	-
Smoker	1 if smokes $\geq$ 1 cigarette per day, 0 otherwise	.310	.462	-	-
Alqprud	1 if prudent alcohol drinker(see text) , 0 otherwise	.460	.498	-	-
Exercise	1 if exercise in last two weeks, 0 otherwise	.463	.499	-	-
Social class					
regsc1s	1 if social class 1 or student,0 otherwise	.063	.243	.067	.250
regsc2	1 if social class 2, 0 otherwise	.242	.429	.269	.443
regsc3a	1 if social class 3 manual, 0 otherwise	.346	.476	.333	.471
regsc4	1 if social class 4, 0 otherwise	.157	.364	.151	.358
regsc5n	1 if social class 5, 0 otherwise	.051	.220	.052	.223
Education					
lhqdg	1 if University Degree, 0 otherwise	.140	.347	.165	.371
lhqhnd	1 if HND or equivalent, 0 otherwise	.124	.330	.142	.349
lhqA	1 if A level or equivalent, 0 otherwise	.049	.217	.039	.194
lhqnone	1 if no formal qualifications,0 otherwise	.447	.497	.424	.494
lhqoth	1 if vocational/professional qualifications, 0 otherwise	.046	.209	.031	.173
Marital Status					
widow	1 if widowed, 0 otherwise	.059	.236	.097	.296
divorce	1 if divorced, 0 otherwise	.036	.187	.059	.236
seprd	1 if separated, 0 otherwise	.017	.128	.021	.144
single	1 if single, 0 otherwise	.152	.359	.094	.292
Work					
part	1 if employed part-time, 0 otherwise	.146	.354	-	-
sick	1 if absent from work due to illness, 0 otherwise	.017	.129	-	-
retd	1 if retired, 0 otherwise	.155	.362	-	-
stdnt	1 if student, 0 otherwise	.007	.080	-	-
keephse	1 if keep house, 0 otherwise	.149	.356	-	-
unemp	1 if unemployed, 0 otherwise	.044	.204	-	-
wkshft	1 if work shifts, 0 otherwise	.083	.277	.077	.267
Ethnic Group					
ethipb	1 if 1 if Indian sub-continent, 0 otherwise	.010	.102	.010	.102
ethbawi	1 if black African/West Indian, 0 otherwise	.007	.082	.007	.082
ethothnw	1 if other non-white, 0 otherwise	.006	.075	.006	.075
Type of Area					
rural	1 if live in rural area, 0 otherwise	.223	.416	.186	.389
suburb	1 if live in suburb, 0 otherwise	.470	.499	.544	.498
Region					
wales	1 if live in Wales, 0 otherwise	.058	.233	-	-
north	1 if live in north, 0 otherwise	.058	.234	-	-
nwest	1 if live in north west, 0 otherwise	.126	.331	-	-
yorks	1 if live in Yorkshire, 0 otherwise	.089	.284	-	-
wmids	1 if live in West Midlands, 0 otherwise	.087	.283	-	-
emids	1 if live in East Midlands, 0 otherwise	.079	.270	-	-
anglia	1 if live in East Anglia, 0 otherwise	.040	.196	-	-
swest	1 if live in South West, 0 otherwise	.077	.267	-	-
london	1 if live in London, 0 otherwise	.092	.288	-	-
scot	1 if live in Scotland, 0 otherwise	.106	.308	-	-
Physical					
male	1 if male, 0 otherwise	.435	.496	.435	.496
height	Height in inches	66.12	3.84	66.06	3.86
age	Age in years	43.96	15.59	50.95	15.58
age2	age^2/100	21.76	14.67	28.38	16.80
age3	age^3/10000	11.82	11.49	17.05	14.73
age4	age^4/1000000	6.89	8.67	10.88	12.28
Tenure					
housown	1 if own house, 0 otherwise	.876	.330	-	-
Household					
nhouse	Number of other people in household	2.16	1.39	-	-
othersmo	1 if other smokers in household, 0 otherwise	.391	.488	-	-
Parental					
mothsmo	1 if only mother smoked, 0 otherwise	.055	.229	-	-
fathsmo	1 if only father smoked, 0 otherwise	.474	.499	-	-
bothsmo	1 if both parents smoked, 0 otherwise	.354	.478	-	-
alpa	Father) non to heavy drinker(0-4)	1.94	1.15	-	-
alma	Mother non to heavy drinker(0-4)	1.19	.993	-	-

Table 6: Selected HALS1 Variable means by sub-sample defined by number of *a priori* 'healthy' behaviours

	FULL $N = 4611$	0 $N = 130 (Exp = 54)$	1/2 $N = 1575 (Exp = 1603)$	3/4 $N = 2469 (Exp = 2677)$	5 $N = 437 (Exp = 277)$
Health and Lifestyle					
Health	.754	.615	.680	.787	.881
Sleepgd	.624	-	-	-	-
Breakfast	.655	-	-	-	-
Smoker	.310	-	-	-	-
Alqprud	.460	-	-	-	-
Exercise	.463	-	-	-	-
Social class					
regsc1s	.063	.023	.025	.076	.140
regsc2	.242	.108	.186	.271	.323
regsc3a	.346	.523	.392	.322	.261
regsc4	.157	.208	.206	.134	.098
regsc5n	.051	.100	.069	.043	.021
Education					
lhqdg	.140	.062	.079	.163	.249
lhqhnd	.124	.092	.107	.131	.156
lhqA	.049	.038	.039	.053	.071
lhqnone	.447	.600	.571	.391	.272
lhqoth	.046	.077	.043	.046	.041
Work					
part	.146	.108	.124	.151	.215
sick	.017	.069	.027	.011	.000
retd	.155	.092	.167	.164	.080
stdnt	.007	0.00	.001	.009	.016
keephse	.149	.146	.168	.140	.130
unemp	.044	.085	.062	.036	.011
wkshft	.083	.131	.098	.077	.053
Physical					
Male	.435	.500	.417	.446	.414
Height	66.12	66.19	65.67	66.36	66.34
Age	43.96	43.96	45.41	43.74	39.99
Tenure					
housown	.876	.908	.897	.866	.847
Household					
nhouse	2.16	2.10	2.14	2.14	2.30
othersmo	.391	.623	.486	.336	.289
Parental					
mothsmo	.055	.085	.054	.053	.067
fathsmo	.474	.423	.477	.480	.446
bothsmo	.354	.408	.358	.349	.352
alpa	1.94	2.15	1.94	1.92	1.96
alma	1.19	1.08	1.03	1.10	1.30

Table 7: Multivariate Probit coefficient estimates for reduced form lifestyle equations(Recursive model(R=50))Standard errors in parentheses

$N = 4611$	SLEEPGD	BREAKFAST	CIGARETTES	ALCOHOL	EXERCISE
soc class					
regsc1s	.22(.10)*	.12(.11)	-.20(.12)	.04(.09)	.19(.11)*
regsc2	.08(.07)	-.10(.07)	-.04(.08)	-.04(.06)	-.03(.07)
regsc3a	-.06(.06)	-.19(.07)*	.21(.07)*	-.16(.06)*	-.10(.07)
regsc4	-.03(.07)	-.23(.08)*	.19(.08)*	-.21(.07)*	-.22(.08)*
regsc5n	-.03(.10)	-.17(.11)	.25(.11)*	-.44(.10)*	-.26(.11)*
education					
lhqdg	-.12(.08)	.20(.08)*	-.21(.09)*	.08(.07)	.07(.08)
lhqhnd	-.04(.07)	-.15(.07)*	-.12(.08)	.01(.06)	-.05(.07)
lhqA	-.05(.10)	.08(.11)	.02(.11)	-.06(.10)	-.04(.10)
lhqnone	-.05(.06)	-.30(.06)*	.14(.06)*	-.14(.05)*	-.24(.06)*
lhqoth	-.003(.10)	-.19(.11)	.17(.11)	-.14(.10)	.04(.11)
marital					
widow	-.13(.09)	-.15(.10)	.21(.11)*	-.07(.09)	.01(.10)
divorce	-.02(.11)	-.13(.11)	.49(.11)*	.03(.10)	.33(.11)*
seprd	-.31(.16)*	-.36(.18)*	.58(.16)*	.04(.16)	.14(.16)
single	-.16(.09)*	-.04(.09)	.22(.09)*	-.32(.08)*	.04(.09)
work					
part	.19(.07)*	.32(.07)*	-.10(.07)	.001(.06)	.05(.07)
sick	-.41(.16)*	-.29(.16)	.44(.16)*	-.85(.18)*	-.41(.18)*
retd	.15(.10)	.51(.11)*	.15(.11)	-.08(.09)	.06(.11)
stdnt	.26(.30)	.33(.30)	-.25(.29)	-.01(.22)	.06(.31)
keephse	-.01(.07)	.10(.07)	.01(.08)	-.22(.06)*	-.18(.07)*
unemp	-.01(.10)	-.05(.10)	.41(.10)*	-.19(.09)*	-.15(.10)
wkshft	-.34(.07)*	-.19(.07)*	.23(.08)*	-.11(.07)	.06(.08)
ethnic					
ethipb	-.11(.20)	-.49(.21)*	.17(.20)	-.98(.27)*	-.55(.24)*
ethbawi	-.39(.27)	-.91(.27)*	-.01(.33)	-.70(.31)*	.16(.25)
ethothnw	-.09(.33)	-.39(.27)	.64(.27)*	-.14(.31)	-.27(.30)
area					
rural	.16(.06)*	.07(.06)	-.06(.06)	.15(.05)*	.03(.06)
suburb	.02(.05)	.10(.05)*	-.01(.05)	.12(.04)*	.04(.05)
region					
wales	-.05(.10)	.02(.10)	.22(.10)*	-.01(.09)	-.10(.10)
north	.04(.10)	.03(.10)	.36(.10)*	-.04(.08)	-.01(.10)
nwest	.10(.07)	.10(.07)	.27(.08)*	-.02(.06)	-.05(.07)
yorks	.05(.08)	-.02(.08)	.13(.09)	-.04(.07)	.02(.08)
wmids	-.10(.08)	-.04(.08)	.18(.09)*	-.08(.07)	-.28(.08)*
emids	-.06(.08)	-.02(.09)	.13(.09)	-.01(.08)	-.02(.08)
anglia	-.01(.11)	-.03(.12)	.19(.12)	.08(.10)	.08(.11)
swest	-.08(.08)	.03(.09)	.06(.09)	-.02(.07)	-.09(.08)
london	-.06(.08)	.01(.09)	.27(.09)*	.03(.07)	-.13(.08)
scot	-.13(.08)	.23(.08)*	.33(.08)*	-.14(.07)*	-.07(.08)
physical					
male	.03(.06)	-.01(.07)	.08(.07)	.09(.06)	.03(.07)
height	-.01(.01)	.01(.01)	-.01(.01)	.002(.01)	.002(.01)
age	-.19(.12)	-.20(.11)	-.09(.14)	-.20(.11)	-.13(.13)
age2	.64(.39)	.73(.38)	.35(.48)	.64(.37)	.36(.43)
age3	-.95(.55)	-1.03(.54)	-.50(.69)	-.83(.52)	-.48(.62)
age4	.49(.28)	.50(.27)	.22(.36)	.38(.26)	.23(.32)
wealth					
housown	-.12(.10)	.12(.10)	.28(.10)*	-.04(.09)	-.10(.10)
household					
nhouse	.01(.02)	-.02(.02)	-.03(.02)	.01(.02)	.01(.02)
othersmo	-.11(.04)*	-.26(.04)*	.66(.05)*	-.08(.04)*	-.04(.04)
parental					
mothsmo	.04(.11)	-.14(.11)	.39(.11)*	.04(.09)	.05(.11)
fathsmo	.10(.07)	-.01(.07)	.17(.08)*	.02(.06)	.01(.07)
bothsmo	.09(.07)	-.23(.08)*	.30(.08)*	.03(.06)	.03(.07)
alpa	-.02(.02)	-.03(.02)	.02(.02)	.01(.02)	.01(.02)
alma	.01(.02)	-.02(.02)	.02(.02)	.10(.02)*	.03(.02)
_cons	3.01(1.31)*	1.67(1.29)	-.34(1.51)*	2.04(1.25)	2.11(1.39)

Table 8: Coefficient estimates for alternative estimators of SAH model(Standard errors in parentheses-* indicates statistical significance at 5% level or below.)

$N = 4611$	MVP(R=50)	Exogenous	Excluded
Lifestyle			
Sleepgd	-.01(.23)	.16(.04)*	-
Breakfast	.16(.21)	-.03(.05)	-
cigarettes	-.53(.16)*	-.34(.05)*	-
alcohol	1.25(.10)*	.09(.04)*	-
exercise	-.14(.20)	.14(.05)*	-
soc class			
regsc1s	.22(.10)*	.33(.12)*	.35(.12)*
regsc2	.07(.06)	.09(.07)	.10(.07)
regsc3a	.06(.06)	-.03(.07)	-.07(.07)
regsc4	.02(.07)	-.11(.08)	-.16(.08)*
regsc5n	.01(.09)	-.18(.10)	-.24(.10)*
education			
lhqdg	-.10(.07)	-.04(.08)	-.01(.08)
lhqhnd	-.04(.06)	-.01(.08)	-.01(.08)
lhqA	.01(.11)	-.01(.12)	.01(.12)
lhqnone	-.18(.05)*	-.32(.06)*	-.36(.06)*
lhqoth	-.03(.11)	-.05(.13)	-.07(.13)
marital			
widow	-.04(.07)	-.09(.08)	-.11(.08)
divorce	-.07(.08)	-.13(.09)	-.18(.09)*
seprd	-.01(.12)	-.04(.15)	-.11(.14)
single	.05(.07)	-.05(.08)	-.10(.08)
work			
wkshft	.25(.07)*	.28(.09)*	.25(.09)*
ethnic			
ethipb	-.03(.22)	-.51(.19)*	-.54(.19)*
ethbawi	.18(.28)	-.23(.24)	-.22(.24)
ethothnw	-.11(.22)	-.22(.27)	-.27(.27)
area			
rural	.06(.06)	.13(.06)*	.16(.06)*
suburb	.02(.04)	.05(.05)	.06(.05)
physical			
male	-.14(.06)*	-.17(.06)*	-.17(.06)*
height	.001(.01)	.01(.01)	.01(.01)
age	.16(.14)	.53(.15)*	.53(.15)*
age2	-.46(.41)	-1.50(.45)*	-1.52(.44)*
age3	.53(.51)	1.77(.54)*	1.80(.54)*
age4	-.22(.22)	-.75(.24)*	-.76(.24)*
_cons	-1.56(1.92)	-6.33(2.00)*	-6.21(1.98)*