

Estimating a Structural Model of Causal Paths between Health and Socio-Economic Status: Evidence from European Older Workforce Surveys¹

Sandra Cavaco*, Jean-Michel Etienne** and Ali Skalli***

April 2006

First draft – Please do not quote

Abstract:

We estimate a structural model describing a variety of dynamic causal paths that are likely to relate different health dimensions to different socio-economic measures. We use Asymptotic Least Squares (ALS) as a means of simultaneously testing such causal mechanisms between health status and socio-economic status. The ALS method provides an appropriate framework to account for important issues such as simultaneity, the existence of intermediate factors and the effect of confounding mechanisms. We use data from the SOCIOLD survey where the targeted population is the older workforce (above 50 years old) from six EU countries; namely Denmark, Finland, France, Greece, the Netherlands and the United-Kingdom. The results enable to disentangle the significant associations of SES components and health conditions. Thus, individuals' health investments and behaviours, i.e. life style factors, have a positive significant effect on individuals' socio-economic history but not on health history. Moreover, socio-economic history turns out to be at the same time a function of individuals' propensity to invest in health and a determinant of current health status but not of current socio-economic status.

JEL Classification: I12, I18, C31

Keywords: Health status, socio-economic status, causal paths, Asymptotic Least Squares

* LEM, Université Panthéon-Assas (Paris 2), 92, rue d'Assas – F-75271 Paris Cedex 06 and CCP, Aarhus School of Business, Denmark .

** ERMES, FRE 2887 CNRS, Université Panthéon-Assas (Paris 2), 12, Place du Panthéon – F-75231 Paris Cedex 05.

*** LEM, Université Panthéon-Assas (Paris 2), 92, rue d'Assas – F-75271 Paris Cedex 06.

¹ We are grateful to Tor Eriksson, Ioannis Theodossiou as well as to participants to the 5th IHEA conference in Barcelona and to the departmental seminar at the Aarhus School of Business for valuable discussions and comments. Financial support from the European Commission under grant QLK6-CT-2002-02292 for the SOCIOLD project is also gratefully acknowledged. The usual disclaimer applies.

1. Introduction

During the last two decades, a fairly large literature has built up suggesting that socio-economic status (SES) has important implications for health.² Whilst few researchers would argue against such a positive relationship both within and across countries (see Adler *et al.*, 1994; van Doorslaer *et al.*, 1997), the main direction of causality between income and health is open to more debate (Smith, 1999; Benzeval *et al.*, 2000; Case *et al.*, 2002; Adams *et al.*, 2003; Meer *et al.*, 2003). On the one hand, as noted by Deaton and Paxton (1998), “There is a well-documented but poorly understood gradient linking socio-economic status to a wide range of health outcomes”. On the other hand, policy makers need a clear picture of the mechanisms underlying the social health gradient to better target the reduction of health inequalities among citizens.

The demand for health model developed by Grossman (1972) assumes that individuals inherit an initial stock of health, which depreciates with age and increases with health investments. The stock of health at a given point of time is the cumulative outcome of an entire history of past resources, past health behaviours and past consumption. The occurrence of a disability can be the result of shocks (life events, accidents, etc.) but it can also result from a gradual process of health deterioration.

Most of the existing literature relies implicitly or explicitly on Grossman’s model of health accumulation, which provides an insight on how health is formed not only from investments in medical care but also from health-related behaviours. One corollary of this model is that if socio-economic differences determine the propensity to invest in medical care and/or lifestyle behaviours, then they should influence individuals’ health as well. Hence, the main issue in empirical studies is that of whether socio-economic differences affect the various dimensions of individual health status. Examples of such analyses are those by Hitiris and Posnett (1992) and Rhum (2000) who both use state-level data and those by Martelin (1994), Everson *et al.*, (2002) and O’Reilly (2002) who rely on individual data. The evidence suggests that social conditions are important determinants of health, even more important than access to health care *per se* (Pincus *et al.*, 1998). Socio-economic factors that were found to affect health inequalities are income (Ecob and Davey Smith, 1999; Soobadeer and Le Clere, 1999; Fiscella and Franks, 2000; Deaton and Paxton, 2001), occupational status indicators (Or, 2000), education (van Rossum *et al.*, 2000; Everson *et al.*, 2002), genetic endowments (Smith

² See Smith (1999), Goldman (2001) and Skalli *et al.* (2006) for comprehensive literature reviews.

and Kington, 1997), lifestyle factors (Or, 2000; Osler *et al.*; 2002, Sturm and Gresenz, 2002; Contoyannis and Jones, 2004) and past demographic history (Grundy and Holt, 2000).

Recent studies have also focused on the importance of early childhood environment on adults' health and labour market outcomes (Case *et al.*, 2002; Currie and Hyson, 1999; Currie and Stabile, 2003; Case *et al.*, 2005). Childhood socio-economic conditions have been found to have an impact on adults' health, independent of their socioeconomic conditions (Blane *et al.*, 1993; Kuh *et al.*, 1997; Reynolds & Ross, 1998). Thus, early childhood conditions have a significant direct effect on the rate of health depreciation as well as on employment statuses over the life cycle. For instance, Lindeboom *et al.* (2005) show that early childhood factors influence both the rate of accidents and the probability of getting a disability. Therefore, both early childhood factors, which are out of individuals' control, and the sequential decision process they have adopted throughout their life cycle determine individuals' health and socio-economic status at later stages of their working lives (Grossman, 1972; Smith, 1999).

Thus, the literature suggests that the effect of socio-economic conditions on health status may take various paths. Put differently, a large number of causal mechanisms might be at play (Preston and Taubman, 1994; Smith and Kington, 1997, Smith, 1999) and the crucial issue is that of how to disentangle causal from ecological effects. In addition, these causal paths are either reinforced or weakened by the effects of *intermediate factors*. For instance, while SES influences the propensity to consume medical care, the latter, in turn, is likely to influence health. Likewise, childhood environment might have an impact on school performance which, in turn, might influence future SES and hence, future health. Moreover, there might be *cofounders* at play as well. That is, factors which influence both health and SES. For instance, time discounting and risk aversion behaviours are likely to have an influence on health investments but also on schooling decisions, the latter being important determinants of SES. Besides, time preference influences health-affecting behaviour like smoking, exercising or following dietary restrictions (Fuchs, 1986). Higher rates of time preference lead to lower demand for longevity and less investment in health (Erlich and Chuma, 1990).

Ideally, one would also need to account for reverse causality as health status may influence SES since poor health is likely to reduce earnings and wealth accumulation power (Smith, 1999). More than multiple causal paths, intermediate factors and cofounders, this issue has been the subject of many studies. In general, however, reverse causality is accounted for by modelling one of the two pathways and by controlling for the endogeneity of the causal (right-hand side) variable. One example of such studies is the one by Chapman and Hariharan

(1994) who attempt to control for the endogeneity of income in a health equation, using information on previous health. An alternative approach consists in using longitudinal data while controlling for individual fixed effects (Smith and Kington, 1997). However, genetically predetermined diseases, environmental factors that are not related to fixed individual traits are also likely to have an effect on SES variables. Thus, controlling for fixed effects does alleviate some endogeneity, but does not necessarily eliminate it.

Besides the possibility they offer to control for individual heterogeneity, panel data also have a number of features that are very helpful given the complexity of the association between health and SES. Adams *et al.* (2003) for instance consider a structural model describing a variety of potential causal paths between individual health history, health events, tastes and behaviours, socio-economic history and socio-economic events. Then, using longitudinal data of American elderly people, they conduct Granger-type causality tests to discriminate between actual causal effects and ecological ones. They come to the conclusion that some of the supposedly possible paths play actually no role.

To tackle these issues, we adopt an analogous approach in this paper. We depart from a structural model in order to identify the association between health and SES that is very similar to Adams *et al.*'s (2003). However in contrast to them, we use cross-section data of the older workforce (50 and older) from 6 EU countries and estimate our structural model using Asymptotic Least Squares (ALS). This way, we are able to test for the validity of each of the underlying causal paths postulated by the model, controlling for selection, reverse causality and simultaneity.

The paper is organised as follows. Section 2 describes the data as well as the empirical set-up. Section 3 discusses the results and section 4 concludes the paper.

2. Data and Empirical Set-up

2.1. Data

We use the so-called SOCIOLD survey which comprises 6 country-specific cross-sectional datasets, covering Denmark, Finland, France, Greece, the Netherlands and the United-Kingdom. The questionnaire has been designed by the partners of the EU-funded SOCIOLD research project and the survey has been conducted in 2004 by interview private companies in the different countries. Interviews have been carried out via internet among the older

workforce (users 50 or older), the targeted number of exploitable questionnaires being 1000 in each country.

The questionnaire was designed in order to address the following broad questions at the centre of the SOCIOLD project; namely:

- How does individual socio-economic and/or occupational status affect the physical and mental health and sense of well being of older individuals of working age?
- How does individual socio-economic and/or occupational status affect the ability of older workers to participate in the labour market?

The resulting data provide new information about a variety of dimensions of both health and socio-economic status of the older workforce. These include objective as well as subjective measures of respondents' health both at the time of interview and in the past, thus a detailed description of both current health status and health history is available. Likewise, SES is described through a large number of indicators, including household income, education, employment status, occupational status. In addition, the questionnaire gathers information about respondents' socio-economic family background as well as about their own socio-economic history. Moreover, for a number of life style indicators (smoking, alcohol consumption, dietary habits, physical activity, etc.) respondents are asked to describe their evolution over time. Eventually, a large number of childhood environment and living conditions indicators is collected, including housing conditions, school attendance, social and economic environment. Detailed descriptive statistics are displayed in Appendix 2, Table A.1.

These data provide us relevant information to estimate our structural model, except perhaps, for accurate genetic endowments or intra-uterine information which are unfortunately missing. Thus, they allow us to estimate the model structure described in the next section, with a view to testing for potential causal paths between more specifically Socio-Economic Status Events (SESE), Health Events (HE), Socio-Economic Status History (SESH), Health History (HH), and Taste and Behaviours (TB).

2.2. The Model

We depart from the model structure suggested by Adams *et al.* (2003) and which is summarised in Fig. 1 below. This figure highlights the main causal paths between health and SES that are likely to be at play. First, genetic factors might influence individuals' tastes and hence, their health accumulation behaviour either in terms of health care consumption or in

terms of life style attitudes. But, as a part of individuals' initial endowments, they are also likely to condition their health history. Second, the figure also states that individual tastes and behaviours are also potential determinants of both her/his socio-economic and health history. For instance, risk and time preferences might simultaneously determine individuals' willingness to invest both in health and in education. Of course, the health events one experiences might result either from her/his health history, from her/his past SES or from both. Poor health during childhood is likely to favour the occurrence of chronic diseases during adult life. Likewise, poorly educated people are likely to be less aware than others of the importance of hygiene or of preventive medical care. High SES households have a better knowledge about how to maintain and improve their health and they can devote more resources to health investments. However, health and SES history are also likely to cumulatively have a direct influence on current SES. On the one hand, unhealthy individuals may be less able to achieve education, earn and accumulate wealth. On the other hand, a poor SES history might result in both low earnings and wealth accumulation and then in relatively low current income.

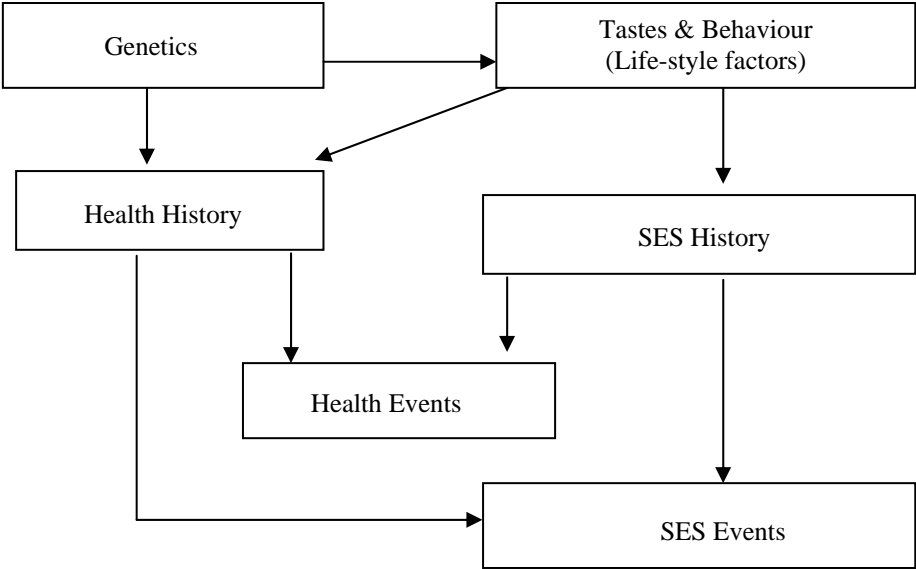


Fig. 1. Causal paths between socio-economic status and health status

We model these causal paths as a system of five simultaneous equations which can be written as:

$$\begin{cases} Y_1 = \gamma_{13}Y_3 + \gamma_{14}Y_4 + X_1\beta_1 + u_1 \\ Y_2 = \gamma_{23}Y_3 + \gamma_{24}Y_4 + X_2\beta_2 + u_2 \\ Y_3 = \gamma_{35}Y_5 + X_3\beta_3 + u_3 \\ Y_4 = \gamma_{45}Y_5 + X_4\beta_4 + u_4 \\ Y_5 = X_5\beta_5 + u_5 \end{cases} \quad (1)$$

where Y_j , $j = 1, \dots, 5$, are measures of socio-economic events (SESE), health events (HE), socio-economic history (SESH), health history (HH) and of tastes and behaviours (TB), respectively, X_j , $j = 1, \dots, 5$, are vectors of exogenous explanatory variables, γ_{jh} and β_j , $j, h = 1, \dots, 5$, are parameter vectors of conformable dimensions and u_j , $j = 1, \dots, 5$, are error terms.

Our testing strategy which consists in simultaneously estimating the above model provides the following advantages:

- It simultaneously accounts for each of the causal paths postulated in Fig. 1;
- It properly treats intermediate factors. As an example, SESH appears in our model as being dependent of TB and as a determinant of SESE;
- Cofounders are taken into account in two ways. First, the X vectors might include common exogenous observable variables which are then considered as determinants of more than one endogenous Y variable. Second, unobserved cofounders are implicitly accounted for through the error structure of the simultaneous-equation model. Hence, the distributions of the u error terms are allowed to share common components;
- Reverse causality is explicitly modelled in the sense that SESH is treated as a determinant of HE whereas HH is treated as a determinant of SESE.

The sub-section explicitly describes each equation of the above model.

2.2.1. *Tastes and Behaviours*

Ideally, the tastes and behaviours index should capture as many health-related individual attitudes as possible. Health is a consequence of various inputs including the adoption of healthy behaviours (dietary habits, exercise etc.) and the avoidance of unhealthy ones (smoking, drinking etc.). It is worth noting that not only are such attitudes likely to directly

influence health status, but they also reflect individuals' time preferences and attitudes towards risk. For instance, 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 as smoking and alcohol consumption. In this paper, we use principal components analysis to construct a single index from a large set of health investment indicators. These include the number of preventive medical visits during the last twelve months, the body mass index (BMI)³, smoking history, exercising history, fish, fruits and vegetables, alcohol and water consumption. All these variables have been coded in such a way that the highest value of each indicator reflects a higher propensity to invest in health. The first principal component is used to construct the composite index of the individuals' tastes and lifestyle behaviours, namely Y_5 .

The resulting variable is thus modelled as a linear function of a set, namely X_5 , of variables describing individuals' genetic endowments and a set of demographic and childhood characteristics which act as control variables. Unfortunately, the data are neither designed to account for genetic factors nor do they contain any explicit information on these. Even though, we attempt to control for these factors by including a variable indicating whether one's father or mother or both have deceased due to some genetic disease.⁴

Demographic characteristics include a gender dummy, age and its square, the number of children, two qualification dummies for upper secondary and higher education (compulsory education is the reference) and three marital status dummies (eg. married, divorced and widowed), single being the omitted group.

Childhood environment is captured through family home characteristics when the respondent was 14 years old. They include the number of rooms in the family home, the number of people who lived in the same family home and a set of dummies indicating whether the respondent had to share her/his bed, whether the house toilet was inside or outside the house, whether there was any crime or vandalism in the area and whether the respondent was brought up in a single parent household.

³ BMI is calculated by the index: weight in kg/(height in m)².

⁴ In case one of, or both their parents have died, respondents had to indicate whether one of the following diseases was the cause of death: cardiovascular/heart disease, cerebro-vascular, respiratory, endocrine, musculo-skeletal, gastro intestinal, genito-urinary, malignant growth, benign growth, diabetes and/or high blood pressure.

Since the principal components approach results in a continuous non-censored variable, the Tastes and Behaviours equation is estimated using ordinary least squares, assuming the u_5 term to be normally distributed.

2.2.2. *The History Equations*

Individuals' history is modelled through a set of two equations, the first one describing their health history, namely Y_4 , and the second one, their socio-economic history, namely Y_3 . We use a probit model to estimate individuals' health history, assuming there exists a latent variable the observed counterpart of which is the self-assessment of individuals of their health status ten years prior to the survey⁵.

Unfortunately, this is the only health history information that is available in the data. The literature on self-reported health indicator argues that subjective measures may lead to substantial under-estimation of the effects of general health. The main criticism relies on the idea that individuals might use health to justify their decision not to work. That is, since health is one of the few legitimate reasons to be out of the labour force, individuals who face poor labour market opportunities rationalize their absence from the labour market by reporting poor health (Lindeboom *et al.*, 2005). Another argument is that relative deprivation, due to income inequality for instance, might induce individuals to assess lower health statuses. In addition, the distributions of subjective measures are in general difficult to compare across countries as they may contain cultural and institutional components.

It should, however, be noted that in this paper's context, health history is measured through a variable measuring how individuals assess their health ten years ago, not their current health. We believe that the above criticisms are less relevant for such a retrospective measure than for current self-assessed health.

Socio-economic history is captured through the total number of months respondents have been unemployed over the last ten years. We thus assume there is a latent socio-economic history measure, namely Y_3 , the observed counterpart of which is the number of past unemployment months. Since a large proportion of individuals reported zero unemployment

⁵ Past self-assessed health is measured in the data through a 5-scale index, the lowest value indicating the poorest health status. The original variable has been translated into a binary indicator with value 1 for individuals reporting good and very good health.

spell, the socio-economic history can be considered as a left-censored dependent variable and be estimated by a tobit model^{6,7}.

Actually in the data, we also have the individuals' employment status ten years prior to the survey (e.g. working, self-employed, unemployed versus non-employed). Our preference goes, however, to the unemployment spell over the last ten years for at least two reasons. First, the available employment status information is rather crude and is thus a very poor measure of socio-economic history. Second, the historical dimension is most likely to be better captured by a measure covering a 10 years span rather than by employment status in a given year.

Given the model structure described in Fig. 1, both health history and socio-economic history variables depend on Tastes and Behaviours index.

However, Fig. 1 also suggests that while genetic factors influence socio-economic history only via their effect on tastes and behaviours, they also have a direct effect on health history. For this reason, they enter the left-hand side of the health history equation and not that of the socio-economic history equation.

Besides, both equations include a set of demographic control variables. Actually, since the history equations relate to the ten years prior to the survey, only non time-varying (gender, education and the number of children) demographic variables are controlled for.

2.2.3. The Events Equations

The model also includes two further equations describing individuals' health and socio-economic events, respectively. It is assumed that there is a latent variable, namely Y_2 , measuring respondents' current health events, the observed counterpart of which is current health status. Actually, though the data contain information on subjective and objective measures of health, we have chosen to resort to the latter measure.

We construct an ADL-score as our health events variable. More specifically, individuals are asked whether, due to chronic physical health conditions, they need help with a number of activities of daily living. They thus indicate how difficult it is for them to bath or dress, to walk one block, to walk several blocks, to walk more than one mile, to bend, kneel or stoop,

⁶ See descriptive statistics in Table A.1 in Appendix 2.

⁷ Although we did not know unfortunately to what number of unemployment spells a given number of months corresponds to, the socio-economic history equation has also been alternatively estimated by a duration model. Estimation, however, did not result in any significant qualitative change in the results.

to climb one flight of stairs, to climb several flights of stairs, to lift or carry groceries, etc. Taking into account all the items, the total difficulty is stated on a 10-point scale, 10, indicating the highest difficulty. As the health events index is a count dependent variable, we use a Poisson regression model.

Finally, we assume there is a latent variable, namely Y_1 , measuring individuals' socio-economic events, the observed counterpart of which is current household income. Unfortunately, household income is provided in the data in the form of 13 income intervals, which makes cross-country comparisons very difficult to perform. Our strategy consisted in constructing a 5-interval classification such that each interval contains 20% of the sample. This way, the resulting variable indicates for each country the quintile to which individuals' household income belongs. Thus, it actually reflects relative, not absolute household income and is hence comparable across countries. It implies that socio-economic events are estimated by an ordered probit.

According to Fig. 1, the main determinants of both health events and socio-economic events are health history as well as socio-economic history. But these equations include the set of demographic characteristics as well, a gender dummy, the number of children, age and its square, the two qualification dummies and the three marital status dummies. These variables are thus common to the two exogenous vectors, respectively X_1 and X_2 . Actually, the only difference between them is that the latter includes a further dummy indicating whether the respondent holds a private health insurance contract in addition to public health insurance. The idea is that this might widen individual's access to health care or access to better quality care and thus might have an influence on current health status.

2.2.4. Identification

The simultaneous equations model in (1) explicitly accounts for endogeneity of both health history and socio-economic history as well as of individuals' tastes and behaviours. Fig. 1 also highlights that the only variable that could be considered as completely exogenous is that related to genetic endowments. However, since it enters the Health History as well as the Tastes and Behaviours equations, it cannot allow to identify properly the whole set of parameters of the model. In our model, genetic factors are assumed to be directly correlated to health history, not only via their effect on tastes and behaviours. They are thus indirectly correlated to health events and to socio-economic events since health history is a determinant of these variables.

Rather, we resort to other variables which could be considered as valid instruments. To identify parameters γ_{35} and γ_{45} respectively in the health history and in the socio-economic history equations, we include as instrument, in the X_5 vector entering the Tastes and Behaviours equation, an indicator of individuals' time preference. This variable highlights specifically individuals' time preference regarding their behaviour towards health. Time preference refers to the rate at which people are willing to trade current benefit for future benefit. A higher rate of time preference will, *ceteris paribus*, lead to less investment in exercise, dietary restrictions etc. Indeed, we do believe that such a time preference indicator is likely to be correlated with health history and events as well as with socio-economic history and events, but only via its effect on individuals' tastes and behaviours. This variable could thus be legitimately excluded from the four first equations of model (1). More precisely, it is measured from the following question:

Suppose that you were asked to stop smoking and in exchange for that you would be guaranteed that you get an extra period of life as an active person in reasonably good health. How long would the minimum of additional life period have to be for you to accept the offer?

One additional year,

Two additional years,

3-4 additional years,

5-6 additional years,

More than 6 additional years,

I would not accept the offer, irrespective of how long the period of additional years offered to me would be.

Individuals responding 'One additional year' have been assigned value 1 while the others have been assigned zero. The latter are considered as having a higher rate of time preference.

To identify parameters γ_{14} and γ_{24} in the health events and in the socio-economic events equations, we include as instrument, in the X_4 vector entering the health history equation, an indicator of individuals' initial health endowments. This variable is constructed from the following question:

When you were less than 12 years of age, did you ever miss one month or more of school for health reasons?

Yes,

No,

The main idea here is that initial health endowments are unlikely to affect health or socio-economic events if they have no influence on individuals' health history.

Finally, to identify parameters γ_{13} and γ_{23} in the health events and in the socio-economic events equations, we include as instrument, in the X_3 vector entering the socio-economic history equation, an indicator of individuals' perception on the role luck plays in life. To be more specific, respondents are asked the following question:

Do you believe that in the long run people get the respect they deserve in this world?

Yes,

No,

Our assumption is that individuals who believe that the society they live in values effort and hard work are likely to have done as much as they could to improve their socio-economic status throughout their working lives and that both their current health and socio-economic statuses are influenced by the effect of such behaviour on their socio-economic history. Put differently, one's health and/or socio-economic statuses in the long run result from the cumulative effect of the efforts they have exerted throughout her/his life, not simply from short run attitudes.

2.3. Estimation Method

Given the data and the way the endogenous $Y_j, j=1, \dots, 5$, variables are measured, the system in Eq. (1) corresponds to a simultaneous equation model mixing a linear model (taste and behaviours) with a probit model (health history), a tobit model (socio-economic history), a *Poisson* regression model (health events) and an ordered probit model (socio-economic history).

To estimate such model structure, we resort to the non-linear Asymptotic Least Squares (ALS) method which has been developed by Gouriéroux, Monfort and Trognon (1985)⁸ as an extension of so-called Amemiya's M-estimation method. As far as we know, though it is rather widely used in other areas of economics (e.g. Crépon *et al.*, 1998; Duguet and Kabla, 1998; Duguet and Greenan, 1999; Galia and Legros, 2004), the ALS method has never been used in the health inequality literature. As Appendix 1 gives an overview of the method and highlights its main properties, we briefly motivate our choice and describe the main steps of the empirical set-up.

As can be seen from Appendix 1, the ALS method yields Minimum Distance Estimators and is, by construction, suitable to estimate simultaneous-equation models where the dependent

⁸ See also Gouriéroux and Monfort (1996).

variables are not of the same type as is the case of system in Eq. (1). Of course, an alternative strategy would consist in resorting to Generalized Method of Moments (GMM) estimators. Our preference, however, goes to ALS rather than GMM for at least two reasons. First, though both methods imply a two-step estimation procedure, GMM requires the use of the whole data at each step whereas ALS's second step simply consists in retrieving the structural model parameters only from auxiliary parameters estimated in the first step. The ALS method is therefore less costly than GMM in terms of computation. The second reason is that the ALS method can be easily extended to handling more complex simultaneous-equation models. It thus provides a unified and tractable framework for the estimation of generalized limited dependent variables systems (Crepon *et al.*, 1998).

Each of the two ALS estimation steps is thoroughly described in Appendix 1. The first one consists in estimating the reduced-form model parameters using the appropriate likelihood maximisation methods (M-estimation). The reduced-form parameters being associated to the structural model parameters through a set of so-called identification constraints, allow in the second step to infer structural-form parameter estimates from the reduced-form parameter estimates obtained in the first step.

The complex structure of the model as well as the asymptotic properties of ALS estimators imply that reliability of the results is crucially dependent on sample size. This is why rather than conducting country-specific analyses, we pool the 6-country completely comparable data sets and conduct our analyses by controlling for country fixed effects in each of the model equations.

3. Results

The estimated coefficients of the structural model are reported in Table 1 below. The most important coefficients are the ones associated with endogenous variables; that is, the γ coefficients. Each of them is indeed associated with a specific causal path and the main purpose of this study is to discriminate between significant causal paths and insignificant ones. However, the assessment of the significant paths requires first to discuss the effect of all the other variables that are controlled for. We thus start by interpreting the explanatory power of the exogenous variables before we turn to interpreting the impact of the endogenous ones.

Table 1: The structural model

Equations	<i>TB</i> (<i>OLS</i>)		<i>HH</i> (<i>Probit</i>)		<i>HE</i> (<i>Poisson model</i>)		<i>SESH</i> (<i>Tobit</i>)		<i>SESE</i> (<i>Ordered probit</i>)	
Variables	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Endogenous variables										
Tastes and Behaviours (TB)			0,034	0,064			-0,387**	0,169		
Health History (HH)					-0,814***	0,053			0,033***	0,009
SES History (SESH)					0,153**	0,076			-0,020	0,027
Genetics										
Health causes of parents' death	-0,013	0,029	-0,086**	0,042						
Childhood environment										
Crime area	-0,191***	0,055								
Number of rooms	0,001	0,003								
Number of persons	-0,017	0,016								
Monoparental family	0,024	0,038								
Toilet inside	0,035	0,037								
Bed sharing	0,029	0,037								
Instruments										
Missed school			-0,373***	0,054						
Lifesmoke	0,739***	0,032								
Responsability							-0,192*	0,106		

*** significant at 1%; ** significant at 5%; * significant at 10%.

Table 1: continued

Equations	<i>TB</i> (<i>OLS</i>)		<i>HH</i> (<i>Probit</i>)		<i>HE</i> (<i>Poisson model</i>)		<i>SESH</i> (<i>Tobit</i>)		<i>SESE</i> (<i>Ordered probit</i>)	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Demographic characteristics										
Female	0,178***	0,027	-0,137***	0,042	0,130***	0,019	0,094	0,109	-0,056*	0,031
Age	-0,050	0,051			0,006	0,019			0,110*	0,059
Age ²	0,056	0,046			0,084**	0,032			-0,117**	0,053
Number of children	0,038	0,028			-0,046	0,042			-0,012	0,027
Upper secondary education	0,096***	0,039	0,152***	0,058	-0,088***	0,023	-0,180	0,155	0,170***	0,045
Higher education	0,302***	0,040	0,197***	0,061	0,172	1,038	-0,273*	0,160	0,506***	0,047
Married	0,005	0,050			0,032	0,037			0,588***	0,056
Divorced	0,014	0,057			-0,017	0,034			-0,025	0,066
Widowed	-0,014	0,074			0,104	0,042			0,035	0,088
Private health insurance					0,019	0,037				
Country fixed effects⁽¹⁾										
France	1,177***	0,049	-0,216**	0,103	0,098*	0,036	0,688***	0,272	0,199***	0,054
Denmark	0,295***	0,054	-0,220***	0,080	-0,095	0,087	1,179***	0,206	-0,195***	0,062
Finland	1,161***	0,064	-0,315***	0,122	-0,463***	0,064	1,108***	0,318	-0,098	0,065
Netherlands	1,003***	0,052	-0,498***	0,098	0,176**	0,101	-0,736***	0,281	0,196***	0,064
United-Kingdom	2,253***	0,052	-0,409***	0,132	-0,057	0,117	2,232***	0,349	0,437***	0,078
Constant	-0,597	1,419	1,096***	0,103	0,222***	0,050	-0,427	0,281	Cuts	Not reported
Number of observations : 5041										

*** significant at 1%; ** significant at 5%; * significant at 10%.

(1) Greece is the omitted country.

An important feature displayed by Table 1 is the statistical significance of the three instrumental variables that are included in the model. This implies that as long as our assumptions regarding the correlation of these variables with the endogenous ones are valid, the γ parameters of the model are properly identified. The signs on these instruments are also economically relevant in the sense that they are exactly the ones one would a priori expect. First in column 1 (TB equation), the positive sign on the coefficient associated with the time preference ‘Lifesmoke’ variable suggests that individuals willing to give up smoking for a 1-year extra period of life have a higher propensity to invest in their health than those who would need a longer extra period of life. Second, column 2 (HH equation) shows that individuals who have missed school for one month or more for health reasons when they were less than 12 years old are less likely to report high past health score. This result highlights that health history is positively correlated with one’s health status during childhood. Thus, there is a temporal persistence in health from early childhood to old age (Smith and Kingston, 1997). Eventually, the negative coefficient associated with the ‘responsibility’ variable suggests that individuals who believe that in the long run people get the respect they deserve have had shorter unemployment durations during the last ten years. Such a result suggests that one’s socio-economic history depends on her/his belief about how society rewards effort.

The other exogenous variables included in the model are country fixed effects as well as a number of demographic characteristics.

Country fixed effects in column 1 (TB equation) suggest that on average, individuals from different countries have different propensities to invest in health. The healthiest lifestyles are observed in Denmark and the riskier behaviours, in Greece. In between, the descending order ranking is the United-Kingdom, France, Finland and the Netherlands. Interestingly, column 2 (HH equation) suggests that the highest past health scores are observed in Greece. The signs on country fixed effects in the three remaining columns are less straightforward to interpret. On average, the Danes, the Finnish and the British have lower health scores than the Greeks, the French and the Dutch.

In contrast, not many demographic characteristics are statistically significant. Neither age nor the number of children nor marital status are significantly correlated with the propensity to invest in health. Probably, this is due to the fact that health behaviours are related to life styles in the long run. The only significant coefficients are those concerning gender and education. They suggest that females have on average less risky behaviours and that the highest is individual’s qualification the highest is her/his propensity to invest in health. Among the

childhood environment variables, none of the housing characteristics seems to have a significant effect on tastes and behaviours; only the incidence of crime and vandalism in the childhood living area negatively influences individuals' propensity to invest in health. The associated coefficient is negative, hence suggesting that, on average, those having spent their childhood in an insecure area are less likely to adopt safe life styles.

The history equations include the gender and education as demographic indicators. We do find a trend for gender. The results for the health history equation show that females as well as the least educated have on average reported lower past health scores. The same variables seem to have less relevance in the socio-economic history equation since most of them are not statistically significant. The only coefficient that is significant at the 10% level is the one associated with the tertiary education dummy, hence suggesting that higher education attendees have on average experienced shorter unemployment durations.

Likewise, very few demographic characteristics seem to influence significantly individuals' current health status. This is the case with the number of children and with marital status, probably due to the sample of elderly people being rather homogeneous with respect to these variables. The coefficient on the private health insurance dummy is not significant either. Again, this might be due to sample homogeneity if a large number of European elderly do hold such contracts. Overall, only the coefficients on gender, age and education show some statistical significance. In line with the gender trend highlighted in the health history equation, females are also in poorer health than males. The relationship between age and ADL-scores is not only positive but convex as well, hence suggesting that health deteriorates with age at increasing speeds. Regarding education, the estimates show that individuals holding an upper secondary education qualification have a better current health status.

Besides, almost all the demographic characteristics significantly influence socio-economic status measured by household income quintiles, except the number of children. Though only significant at the 10% level, there seems to be a gender differential in favour of men. Also, age positively influences the rank-order of individuals in the income distribution, although this influence is decreasing. Likewise, the income rank-order is most likely to increase with qualifications. Eventually, it is also significantly affected by marital status, the married being on average more likely to belong to high quintiles of the household income distribution.

We interpret now the main structural coefficients of model (1). We first note from column 1 (TB equation) that the effect of the genetic factors indicator is not significant. Overall, beside

demographic characteristics and country fixed effects, the only determinants of tastes and behaviours are the crime incidence in the living area during childhood and the 'lifesmoke' time preference indicator.

Interestingly, tastes and behaviours do not have a significant influence on health history, but do impact socio-economic history. The coefficient associated with tastes and behaviours on column 4 (SESH equation) suggests that the higher is individual's propensity to invest in health, the less likely she/he is to experience long unemployment durations. The statistical insignificance of the effect of tastes and behaviours on health history may be linked to the subjective nature of this variable. In fact, individuals could have self-assessed their past health conditionally on their life styles. Besides for instance, a heavy smoker whose health was not excellent ten years earlier might have reported a good past health status if she/he compares her/himself to smokers only, not to the whole population.

Note that column 2 (HH equation) shows that past self-assessed health is also correlated with the health causes of parents' death (genetic factors). The negative sign on the latter variable suggests that those whose parents have died from some genetic diseases are also less likely to report high past health statuses. This effect could be interpreted in terms of inter-generational transmission of health. Summing up, beside demographic variables and country fixed effects, our indicator of individuals' health history depends only on the health causes of parents' death and on own health during childhood. In contrast, individuals' socio-economic history depends on their tastes and behaviours and on their judgement on how the society is likely to recognise and to reward individuals' effort.

Regarding the events equations in columns 3 (HE equation) and 5 (SESE equation), it turns out that not the same effects are at play. Health history is positively correlated with both current health status and current socio-economic status. More specifically, individuals having reported high past health scores are more likely to enjoy good health in the present (i.e. healthy individuals have less difficulties to make daily living activities), but also to belong to the highest quintiles of the income distribution. Socio-economic history on the other hand has a significant effect on health events, not on current socio-economic status. The longer is the unemployment duration one has experienced over the last ten years, the more disabilities and limitations she/he would report at the time of interview. However, it would not have a significant effect on current income level. It is thus worth noting that socio-economic status is measured by household and not individual income. As stated by the added worker hypothesis in labour supply theory, unemployment might, via an income effect, induce other household

members to participate to the labour market in order to maintain current household income at its initial level.

Overall, the results discussed above show that not all the causal paths highlighted in Fig. 1 are at play. Fig. 2 below is a modified version of Fig. 1, where the insignificant causal paths are depicted by dotted lines.

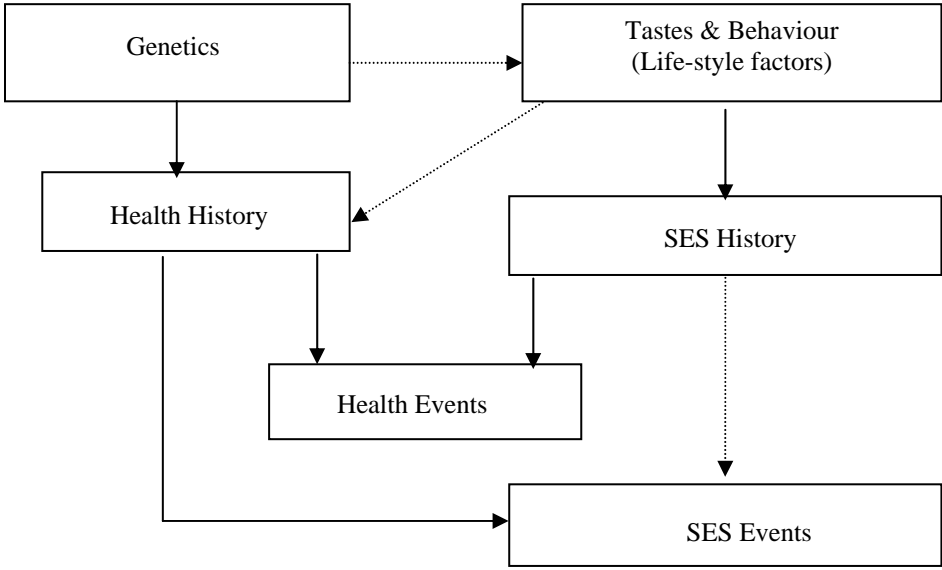


Fig. 2. Significant and insignificant causal paths between socio-economic status and health status

4. Concluding remarks

In this paper, a structural model describing a variety of possible causal paths between health and socio-economic status has been estimated. The model is designed in such a way that it explicitly accounts for the endogeneity of both health and socio-economic statuses and the existence of intermediate factors such as lifestyle behaviours as well as third factors such as time discounting. Therefore, the model controls for reverse causality and simultaneity regarding the health-SES gradient. Based on a rich data set combining samples of old workforce individuals from six European countries, our analysis provides evidence of the high significance of only some causal paths our model structure a priori assumes.

To be more specific, the results suggest that genetic factors influence one’s health history which in turn influences both health and socio-economic status. However, genetics do not have an effect on tastes and behaviours. Individuals’ health investments and attitudes

significantly impact individuals' socio-economic history, not health history. But most interestingly, socio-economic history turns out to be at the same time a function of individuals' propensity to invest in health and a determinant of current health status, not of current socio-economic status.

Furthermore, we may note that, if valid, these results hold for the population sampled in our data, but probably not for the whole population. Some of the causal paths highlighted in our analysis are similar to the ones Adams *et al.* (2003) provide evidence for, based on data on American elderly. However, the results might change for a different specification of the estimated structural model. First, our measure of health does not account for any mental dimension of health. Second, our measure of socio-economic history, based on past unemployment duration, does probably not account for some important socio-economic dimensions. Finally, the causes of parents' death are certainly not sufficiently reliable indicators of genetic factors. This means that these measurement and specification issues should be the subject of further research in order to assess the robustness of the results suggested by our analysis.

References

- Adams P., M. D. Hurd, D. McFadden, A. Merrill and T. Ribeiro (2003) "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status", *Journal of Econometrics*, 112, 3-56.
- Adler N.E., T. Boyce, M.A. Chesney *et al.* (1994) "Socioeconomic Status and Health, the Challenge of the Gradient", *American Psychologist*, 49, 15-24.
- Amemiya T. (1985), *Advanced Econometrics*, Basil Blackwell.
- Benzeval M., J. Taylor and K. Judge (2000) "Evidence on the Relationship between Low Income and Poor Health: Is the Government Doing Enough?", *Fiscal Studies*, 21, 375-399.
- Blane D., D. G. Smith and M. Bartley (1993) "Social Selection: What does it Contribute to Social Class Differences in Health?" *Sociology of Health and Illness*, 15, 1-15.
- Case A., D. Lubotsky and C. Paxson (2002) "Economic Status and Health in Childhood: The Origins of the Gradient", *American Economic Review*, 92, 1308-1334.
- Case A., A. Fertig and C. Paxson (2005) "The Lasting Impact of Childhood Health and Circumstances", *Journal of Health Economics*, 24, 365-389.
- Chapman K. and G. Hariharan (1994) "Controlling for Causality in the Link from Income to Mortality", *Journal of Risk and Uncertainty*, 8, 1, 85-93.
- Contoyannis P. and A. M. Jones (2004) "Socio-economic Status, Health and Lifestyle", *Journal of Health Economics*, 23, 965-995.
- Currie J. and R. Hyson (1999) "Is the Impact of Health Shocks Cushioned by Socioeconomic Status?: The Case of Birth Weight", *American Economic Review*, 89, 2, 245-250.
- Currie J. and M. Stabile (2003) "Socioeconomic Status and Health: Why is the Relationship Stronger for Older Children?", *American Economic Review*, 93, 5, 1813-1823.
- Crépon, B., E. Duguet and J. Mairesse (1998), *Research, innovation, and productivity : An econometric analysis at the firm level*, NBER working paper #6696.
- Deaton, A. and C. Paxson (1998) "Aging the Inequality in Income and Health", *American Economic Review*, 88, 248-253.
- Deaton A. and C. Paxson (2001) "Mortality, Education, Income and Inequality among American cohorts", in D. Wise ed., *Themes in the Economics of Aging*, University of Chicago Press, Chicago.
- Duguet E. and N. Greenan (1999) "Skilled Biased Technological Change at the Firm Level: Econometric Evidence from French Manufacturing", *Les Cahiers de l'Innovation*, cahier n° 99006, University Paris I Panthéon-Sorbonne.
- Duguet E. and I. Kabla (1998) "Appropriation Strategy and the Motivations to Use the Patent System: An Econometric Analysis at the Firm Level", *Annales d'Economie et de Statistiques*, 9/50, 289-328.
- Ecob R. and G. Davey Smith (1999) "Income and Health: What is the Nature of the Relationship?", *Social Science & Medicine*, 48, 693-705.

- Ehrlich I. and H. Chuma (1990) “A Model of the Demand for Longevity and the Value of Life Extension”, *Journal of Political Economics*, 98, 4, 761–782.
- Everson S. A., S. C. Maty, J. W. Lynch and G. A. Kaplan (2002) “Epidemiologic Evidence for the Relation between Socioeconomic Status and Depression, Obesity and Diabetes”, *Journal of Psychosomatic Research*, 53, 891-895.
- Fiscella K. and P. Franks (2000) “Individual Income, Income Inequality, Health and Mortality: What are the Relationships?”, *Health Services research*, 35, 307-318.
- Fuchs V. (1986), *The Health Economy*, Harvard University Press, Cambridge MA.
- Goldman N. (2001) “Social Inequalities in Health - Disentangling the Underlying Mechanisms, in Weinstein M., Hermalin A.I. and Stoto M.A. (eds), *Population Health and Aging, Strengthening the Dialogue between Epidemiology and Demography*. New York : The New York Academy of Sciences.
- Gouriéroux C., A. Monfort and A. Trognon (1985), *Moindres Carrés Asymptotiques*, *Annales de l'INSEE*, 58, 91-122.
- Gouriéroux, C. and A. Monfort (1996), *Statistics and Econometric Models*, Cambridge University Press, Cambridge.
- Grossman, M. (1972) On the concept of health capital and the demand for health. *Journal of Political Economy*, 80, 2, 223-255.
- Grundy E. and G. Holt (2000) “Adult Life Experiences and Health in Early Old Age in Great Britain”, *Social Science and Medicine*, 51, 1061-1074.
- Hitiris T. and J. Posnett (1992) “The Determinants and Effects of Health Expenditure in Developed Countries”, *Journal of Health Economics*, 11, 173-181.
- Kuh D., C. Power, D. Blane and M. Bartley (1997) “Social Pathways between Childhood and Adult Health”, in *A Life Course Approach to Chronic Disease Epidemiology*, edited by D. Kuh and Y. Ben-Shlomo, New York: Oxford University Press.
- Lindeboom M., A. Llana-Noza and B. van der Klaauw (2005) “The Interrelation between Disability and Work and the Role of Health Shocks”, working paper.
- Legros D. and F. Galia (2004) “Research and Development, Innovation, Training, Products’ Quality and Profitability: Econometric Evidence from France”, working paper ERMES, n°04-11.
- Martelin T. (1994) “Mortality by Indicators of Socioeconomic Status among the Finnish Elderly”, *Social Science & Medicine*, 38, 1257-1278.
- Meer J., D. Miller H. and Rosen (2003) “Exploring the Health-Wealth Nexus”, *Journal of Health Economics*, 22, 713-730.
- Or Z. (2000) “Determinants of Health Outcomes in Industrialised Countries: A Pooled, Cross-country, Time-series Analysis”, *OECD Economic Studies*, n°30, 2000/1.
- O’ Reilly D. (2002) “Standard Indicators of Deprivation: Do they Disadvantage Older People?”, *Age and Ageing*, 31, 197-202.

- Osler M., E. Prescott, M. Grønbaek, U. Christensen, P. Due and G. Engholm (2002) "Income Inequality, Individual Income and Mortality in Danish Adults: Analysis of Pooled Data from two Cohort Studies", *British Medical Journal*, 324, 7313-7328.
- Pincus T., R. Esther, A.D. De Walt and L. F. Callahan (1998) "Social Conditions and Self-Management are more Powerful Indicators of Health than Access to Care", *Annals of Internal Medicine*, 129, 5, 406-411.
- Preston, S. and P. Taubman (1994) "Socioeconomic Differences in Adult Mortality and Health Status", In Martin L. and S. Preston eds. *Demography and Aging*, National Academy Press, Washington, 279-318.
- Reynolds J. and C. Ross (1998) "Social Stratification and Health: Education's Benefit beyond Economic Status and Social Origins", *Social Problems*, 45, 221-247.
- Ruhm C. J. (2000) "Are Recessions Good for your Health?", *The Quarterly Journal of Economics*, 617-650.
- Shibuya, K., Hashimoto, H. and Yano, E. (2002) Individual Income, income distribution, and self rated health in Japan: cross sectional analysis of nationally representative sample, *British Medical Journal*, 324.
- Skalli A, E. Johansson and I. Theodossiou (2006), Are the Healthier Wealthier or the Wealthier Healthier? The European Evidence, ETLA, Helsinki.
- Smith J. (1999) "Healthy Bodies and Thick Wallets: the Dual Relation between Health and Economic Status", *Journal of Economic Perspectives*, 13, 155-166.
- Smith, J. P. and Kington R. (1997) "Demographic and Economic Correlates of Health in Old Age", *Demography*, 34, 159-170.
- Soobadeer M-J. and F. LeClere (1999) "Aggregation and the Measurement of Income Inequality: Effects on Morbidity", *Social Science and Medicine*, 48, 733-744.
- Sturm, R. and C. R. Gresenz (2002) "Relations of Income Inequality and Family Income to Chronic Medical Conditions and Mental Health Disorders: National Survey", *British Medical Journal*, 324.
- van Doorslaer E., A. Wagstaff, H. Bleichrodt, S. Calonge, U. Gerdtham and M. Gerfin (1997) "Income-related Inequalities in Health: Some International Comparisons", *Journal of Health Economics*, 16, 93-122.
- van Rossum C. T. M., H. van de Mheen, J. P. Mackenbach and D. E. Grobbee (2000) "Socioeconomic Status and Mortality in Dutch Elderly People - The Rotterdam Study", *European Journal of Public Health*, 10, 255-261.

Appendix 1

Method of estimation: ALS

The key idea of ALS is to estimate reduced form coefficients in each equation of the model separately and then infer from these auxiliary parameters the structural form parameters of the model, using a minimum distance estimator. The intuition is the same as the method of indirect least squares, although ALS is more general. We thus have two estimation issues: first, estimating the reduced form parameters and their joint covariance matrix; then estimating the structural form parameters in a consistent and efficient way. The first problem is solved by interpreting the maximum likelihood estimators of the reduced form equations as specific M-estimators. The second problem comes down to writing explicitly the relationships between the structural form parameters (parameters of interest) and the reduced form parameters (auxiliary parameters).

Reduced form estimation

Our structural model consists of five equations:

$$\begin{cases} Y_1 = \gamma_{13}Y_3 + \gamma_{14}Y_4 + X_1\beta_1 + u_1 \\ Y_2 = \gamma_{23}Y_3 + \gamma_{24}Y_4 + X_2\beta_2 + u_2 \\ Y_3 = \gamma_{35}Y_5 + X_3\beta_3 + u_3 \\ Y_4 = \gamma_{45}Y_5 + X_4\beta_4 + u_4 \\ Y_5 = X_5\beta_5 + u_5 \end{cases} \quad (1)$$

where Y_j , $j = 1, \dots, 5$, are measures of socio-economic events (SESE), health events (HE), socio-economic history (SESH), health history (HH) and of tastes and behaviours (TB), respectively, X_j , $j = 1, \dots, 5$, are vectors of exogenous explanatory variables, γ_{jh} and β_j , $j, h = 1, \dots, 5$, are parameter vectors of conformable dimensions and u_j , $j = 1, \dots, 5$, are random disturbances. The five equations of the model are an ordered probit model, a *Poisson* regression model, a Tobit model, a probit model and a linear model, respectively.

Once the reduced form of the model is written, the estimation problem becomes that of a series of single equations and the estimation method that is the most appropriate for each equation can be applied. One can thus write:

$$\hat{\pi}_k \in \arg \max_{\pi_k} L_k(\pi_k), \quad k = 1, \dots, 5$$

where π_k is the reduced form parameter in equation k and L_k , the corresponding likelihood function. The problem remains to estimate the joint covariance matrix of the $\hat{\gamma}_k$'s. We can solve it by considering that our estimators can also be defined globally by:

$$\hat{\pi} \in \arg \max_{\pi} L(\pi) \text{ with } \pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_5 \end{pmatrix} \text{ and } L(\pi) = \sum_{k=1}^5 L_k(\gamma_k)$$

Maximising L with respect to π gives exactly the maximum likelihood estimators. This property arises from the separability of $L(\pi)$ with respect to the $\hat{\gamma}_k$'s. These estimators can be interpreted as M-estimators. Under the usual regularity conditions for M-estimators:

$$\sqrt{N}(\hat{\pi} - \pi) \xrightarrow{N \rightarrow +\infty} N(0, \Omega)$$

with asymptotic covariance matrix:

$$\Omega = J^{-1} I J^{-1}$$

where

$$I = E_X E_Y \left[\frac{\partial L}{\partial \pi}(\pi_0) \frac{\partial L}{\partial \pi'}(\pi_0) \right], \quad J = E_X E_Y \left[- \frac{\partial^2 L}{\partial \pi \partial \pi'}(\pi_0) \right]$$

where π_0 denotes the true value of parameter π and the expectations are taken with respect to the distributions of the exogenous variables (index X) and of the endogenous variables (index Y).

The matrix I and the matrix J can be estimated by their sample counterparts:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \frac{\partial L_i}{\partial \pi}(\hat{\pi}) \frac{\partial L_i}{\partial \pi'}(\hat{\pi}), \quad \hat{J} = - \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 L_i}{\partial \pi \partial \pi'}(\hat{\pi})$$

where N is the sample size. Note that since the objective function is separable, the off-diagonal terms (the cross derivatives) in \hat{J} (and J) are zero.

Eventually,

$$\hat{\Omega} = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$$

can be used as the estimator of the covariance matrix of the estimated reduced form parameters.

Through $\hat{\Omega}$ (and \hat{I}), the correlations between the perturbations of the different equations of the model are taken into account, without making any specific distributional assumptions on their joint distribution.

The reduced form:

Let us write the reduced form:

$$\begin{cases} Y_1 = X\pi_1 + v_1 \\ Y_2 = X\pi_2 + v_2 \\ Y_3 = X\pi_3 + v_3 \\ Y_4 = X\pi_4 + v_4 \\ Y_5 = X\pi_5 + v_5 \end{cases} \quad (2)$$

where:

$$X_{(N,k)} = \begin{bmatrix} X_1 \\ (N,k_1) \\ X_2 \\ (N,k_2) \\ X_3 \\ (N,k_3) \\ X_4 \\ (N,k_4) \\ X_5 \\ (N,k_5) \end{bmatrix}$$

Define exclusion matrices S_k , $k = 1, \dots, 5$, consisting of 1's and 0's at the appropriate places, such that:

$$XS_k = X_k$$

Then, substituting of (2) into (1) and identifying, we get:

$$\begin{cases} \pi_1 = \gamma_{13}\pi_3 + \gamma_{14}\pi_4 + S_1\beta_1 \\ \pi_2 = \gamma_{23}\pi_3 + \gamma_{24}\pi_4 + S_2\beta_2 \\ \pi_3 = \gamma_{35}\pi_5 + S_3\beta_3 \\ \pi_4 = \gamma_{45}\pi_5 + S_4\beta_4 \\ \pi_5 = S_5\beta_5 \end{cases} \quad (3)$$

Amemiya suggested directly estimating the system (3) using regression methods. Let us use the following notations:

$$\begin{aligned}
\hat{\Pi}_1 &= (\hat{\pi}_3, \hat{\pi}_4, S_1) & \theta'_1 &= (\gamma_{13}, \gamma_{14}, \beta_1) \\
\hat{\Pi}_2 &= (\hat{\pi}_3, \hat{\pi}_4, S_2) & \theta'_2 &= (\gamma_{13}, \gamma_{14}, \beta_2) \\
\hat{\Pi}_3 &= (\hat{\pi}_5, S_3) & \theta'_3 &= (\gamma_{35}, \beta_3) \\
\hat{\Pi}_4 &= (\hat{\pi}_5, S_4) & \theta'_4 &= (\gamma_{45}, \beta_4) \\
\hat{\Pi}_5 &= (S_5) & \theta'_5 &= (\beta_5)
\end{aligned}$$

and:

$$\eta_k = (\hat{\pi}_k - \pi_k) - \sum_{m \neq k} (\hat{\pi}_m - \pi_m), \quad k = 1, \dots, 5$$

We get:

$$\begin{cases}
\pi_1 = \hat{\Pi}_1 \theta_1 + \eta_1 \\
\pi_2 = \hat{\Pi}_2 \theta_2 + \eta_2 \\
\pi_3 = \hat{\Pi}_3 \theta_3 + \eta_3 \\
\pi_4 = \hat{\Pi}_4 \theta_4 + \eta_4 \\
\pi_5 = \hat{\Pi}_5 \theta_5 + \eta_5
\end{cases}$$

or, in matrix notations:

$$\pi = \hat{\Pi} \theta + \eta.$$

where $\eta = g(\theta, \hat{\Pi})$.

The basic idea of ALS is to use the estimate $\hat{\Pi}$ of π obtained in a first step and compute in a second step an estimate $\hat{\theta}$ of θ such that $g(\hat{\theta}, \hat{\Pi})$ is as 'close to zero' as possible. That is, we solve the program:

$$\hat{\theta} = \arg \min_{\theta} g(\theta, \hat{\Pi})' \Psi^{-1} g(\theta, \hat{\Pi})$$

where Ψ is a metric. Whatever Ψ is, $\hat{\theta}$ is consistent if $\hat{\Pi}$ is consistent, and it is asymptotically efficient for Ψ given by:

$$\Psi^* = V_{\text{asympt}} [g(\theta, \hat{\Pi})] = \frac{\partial g}{\partial \Pi'}(\theta, \hat{\Pi}) \Omega \frac{\partial g'}{\partial \Pi}(\theta, \hat{\Pi})$$

where $\Omega = V_{\text{asympt}} [\hat{\Pi}]$ is the covariance matrix of $\hat{\Pi}$. Since we need an estimate of θ to estimate

the optimal metric Ψ^* , we may estimate θ in two steps:

In the first step, we can use for Ψ the Euclidian metric and estimate θ as:

$$\hat{\theta} = \arg \min_{\theta} g(\theta, \hat{\Pi})' g(\theta, \hat{\Pi})$$

and the asymptotic distribution of $\hat{\theta}$ is then:

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{N \rightarrow +\infty} N(0, \Sigma)$$

with:

$$\Sigma = \left(\frac{\partial g'}{\partial \theta} \frac{\partial g}{\partial \theta'} \right)^{-1} \frac{\partial g'}{\partial \theta} \frac{\partial g}{\partial \Pi'} \Omega \frac{\partial g'}{\partial \Pi} \frac{\partial g}{\partial \theta'} \left(\frac{\partial g'}{\partial \theta} \frac{\partial g}{\partial \theta'} \right)^{-1}$$

Replacing θ , Π and Ω by their consistent estimates $\hat{\theta}$, $\hat{\Pi}$ and $\hat{\Omega}$, we are now able to compute an estimated $\hat{\Psi}^*$ such that:

$$\hat{\Psi}^* = \frac{\partial g}{\partial \Pi'}(\hat{\theta}, \hat{\Pi}) \hat{\Omega} \frac{\partial g'}{\partial \Pi}(\hat{\theta}, \hat{\Pi})$$

In the second step, we can then compute the corresponding optimal ALS estimator θ^* such that:

$$\hat{\theta} = \arg \min_{\theta} g(\theta, \hat{\Pi})' \hat{\Psi}^{*-1} g(\theta, \hat{\Pi})$$

and the asymptotic distribution of θ^* is:

$$\sqrt{N}(\theta^* - \theta) \xrightarrow{N \rightarrow +\infty} N(0, \Sigma^*)$$

where Σ^* can be estimated by:

$$\Sigma^* = \left(\frac{\partial g'}{\partial \theta}(\hat{\theta}, \hat{\Pi}) \hat{\Psi}^{*-1} \frac{\partial g}{\partial \theta'}(\hat{\theta}, \hat{\Pi}) \right)^{-1}.$$

Appendix 2

Table A.1: Descriptive statistics of the variables (pooled data)

Variables	Percentages / Means
Tastes & Behaviours (TB): investment in health	
Fish consumption per week	1.54
Fruit & vegetables consumption per week	7.83
Water glasses per day	6.32
Smoking history index (0-4): 4=never smoke	2.39
Exercise history index: 0=never exercise	2.17
Alcohol glasses per week	6.94
At a least 1 preventive medical visit (%)	63.12
BMI	27.62
Health events (HE): ADL-score (%)	
0	28.09
1	21.48
2	13.09
3	9.36
4	6.11
5	5.40
6	3.69
7	2.54
8	2.42
9	3.27
10	4.54
SES history (SESH)	
Non-censored observations (%)	29.79
Unemployment duration in months over 10 years	29.43
SES events (SESE): household income (%)	
1 st quintile	20.25
2 nd quintile	18.19
3 rd quintile	20.89
4 th quintile	17.91
5 th quintile	22.75
Health history (HH) (%)	
Very good/good self-assessed status	78.14

Table A.1: continued

Variables	Percentages / Means
Demographic characteristics	
Female (%)	50.97
Age	55.52
Single (%)	8.19
Married (%)	72.08
Divorced (%)	13.65
Widowed (%)	6.08
Number of children in the household	0.22
Compulsory education (%)	18.80
Upper secondary education (%)	36.89
Higher education (%)	44.31
Private health insurance contract (%)	26.60
Childhood environment	
Number of rooms	4.65
Number of persons	2.49
Monoparental family (%)	16.88
Toilet inside (%)	72.23
Bed alone (%)	19.18
Crime area during childhood (%)	6.42
Genetics	
Health causes of parents' death (%)	56.01
Missed school for health problems (%)	14.72
'Lifsmoke' time preference indicator (%)	58.39
Responsibility (%)	50.80
Country effects (%)	
Denmark	17.19
Finland	8.97
France	18.95
Greece	18.76
Netherlands	18.51
United Kingdom	17.62
Number of observations	5041