The dynamics of health limitations across Europe: a longitudinal analysis using the European Community Household Panel

Cristina Hernández Quevedo*, Andrew M Jones*, Nigel Rice**
* Department of Economics and Related Studies, University of York
** Centre for Health Economics, University of York

PAPER PREPARED FOR HESG06, YORK (26 – 28 JULY, 2006)

WORK IN PROGRESS.
NOT TO BE QUOTED WITHOUT PERMISSION FROM THE AUTHORS.

Abstract

This paper investigates dynamics, heterogeneity and the effect of socioeconomic characteristics in health limitations within and between the EU-15 Member States. In particular, we are interested in whether and to what extent, socioeconomic characteristics as education, income and job status affect health limitations and how this varies across time and countries for the 8 waves of data available in the European Community Household Panel (ECHP). We focus on the dynamics of two binary measures of health limitations, constructed from the answers to the question: "Are you hampered in your daily activities by any physical or mental health problem, illness or disability?". We specify a dynamic latent variable for health limitations and we estimate both pooled and random effects probit and logit models together with complementary log-log models. Preliminary results show that the pooled probit specification is preferred, according to the Akaike and Bayesian Information criteria.

Keywords: health limitations, dynamic models, panel data

The *European Community Household Panel Users' Database*, version of December 2003, was supplied by Eurostat. This paper derives from the project "The dynamics of income, health and inequality over the lifecycle" (known as ECuity III Project), which is funded in part by the European Commission's Quality of Life and Management of Living Resources programme (contract QLK6-CT-2002-02297). We are grateful to participants at the EPUNet 06 Conference (Barcelona, 2006) and at HEDG Seminars (University of York, 2006) for their comments on a preliminary version of this paper.

1. Introduction

Panel data models provide additional information on the relationship between income and health, allowing us to explore in more detail the direction of causation, which is not possible to identify with cross-sectional data (Benzeval & Judge, 2001).
A dynamic analysis is needed to explore the association between income and health, where time is a very important element to consider, in particular, when the lifetime of the individual is of interest (Blundell & Preston, 1995; Williams & Cookson, 2000). This life-course approach is also interesting when determining health inequalities, although a short-panel with 8 waves as we present here, does not allow to account for a great amount of the life course of the individual. Besides, health can gradually be modified by age, job status, work environment and working conditions together with events that may have an impact on individual's health, hence, regarding individual health as a dynamic concept (Smith, 2004).
There are several studies that consider in depth the relationship between income and health over time (Smith, 2004; Adams et al, 2003; Benzeval & Judge, 2001; Benzeval, Taylor and Judge, 2000). Benzeval & Judge (2001) present a list of publications on this issue that focus on adults' health outcomes (see Table 1 in Benzeval & Judge, 2001). They conclude that the principal direction of causation goes from income to health and that health selection is regarded as unimportant in most studies.
Several limitations have been identified in the studies of the relationship between income and health: 1. the use of occupational class as a proxy for income, 2. the potential presence of reverse causality, 3. the limitations of cross-sectional data when taking into account a dynamic perspective.
The objective of this paper is two-sided: 1. to explore a particular estimation strategy to analyse the relationship between socioeconomic status and health limitations using panel data models in a dynamic setting, 2. to provide some policy implications derived from the cross-country comparison of the estimates.
This paper uses the *European Community Household Panel Users' Database* (ECHP-UDB) to analyse the dynamics of the socioeconomic (SE) gradient in two binary indicators of health limitations across European Union Member States. The ECHP-UDB is a standardised annual longitudinal survey, which provides 8 waves (1994-2001) of comparable micro-data about living conditions in the European Union Member States (EU-15). Our analysis focus on two binary measures of health limitations, constructed from the answers to the question: "Are you hampered in your daily activities by any physical or mental health problem, illness or disability?", included in the ECHP-UDB.
The main aim is to investigate the causal effect of the SE characteristics in health limitations within and between the Member States of the European Union. For that purpose, we exploite the longitudinal nature of the ECHP-UDB. We are interested in whether, and to what extent, SE variables as education, income and job status affect health limitations and how this varies across time and countries included in the ECHP-UDB, considering the individuals by groups of age and sex.
The ECHP is a longitudinal dataset that allows the researcher to explore the differences in the socioeconomic gradient in health across countries and to perform comparative analysis between the different European countries included in the dataset. The panel nature of the dataset allows us to run both pooled and random effects probit models. Taking into account the longitudinal perspective of the data, we are provided with additional information on the dynamics of individual health limitations and income. A long-run perspective gives us useful information for public health policies, if policymakers are interested in the lifetime history of the individual.

Little attention has been devoted to health dynamics in the past. We are interested in studying the causal effect of SE factors in two indicators of health limitations from a dynamic perspective, which provides us with richer information about this relationship than a cross-sectional analysis. Even though policies have been applied to reduce the level of inequalities in health, there is evidence that overall inequalities persist over time (Contoyannis, Jones and Rice, 2004) and a dynamic approach should be taken into account.

The focal points of the dynamic analysis as reported by Contoyannis, Jones and Rice (2004) are the following: contributions of state dependence, heterogeneity and serial correlation, issues that we analyse in this study.

Finally, attention will be devoted to study the existence of health-related attrition and its consequences. Failing to account for attrition leads to misleading estimates of health dynamics and the relationship between health and SE status (Contoyannis, Jones and Rice, 2004).

2. Literature Review

There are several studies that investigate the causal effect of SE characteristics on health. However, it has been argued that this gradient is not well understood (Ettner, 1996; Deaton & Paxon, 1998; Benzeval et al, 2000) mainly due to two issues: 1. the use of occupational class as a proxy for income, which creates some confusion, 2. the possibility of reverse causality, i.e. that poor health can lead to low income as well as vice versa (Benzeval et al, 2000).

The study of the association between health and SE status is an issue of relevance for public health, as policy indications can be derived to reduce health inequalities and improve health in each society (Ettner, 1996; Fritjers et al, 2003).

The previous literature on this association is limited, because most of the studies focus on cross-sectional data, as reported by Fritjers et al (2003) and Benzeval & Judge (2001). In fact, panel data provides further information on dynamics of individual health and income and its impact on inequalities on these periods, as recognised by Contoyannis et al (2004). Besides, it allows to take into account the lifetime history of the individual, which could provide useful information for policymakers interested in that approach (Williams & Cookson, 2000). However, cross-sectional surveys do not allow neither to give much evidence on causal effects (Fritjers et al, 2003) nor in the direction of the causation (Benzeval & Judge, 2001).

Several limitations have been mentioned in the literature related to the type of study we want to pursue: 1.Presence of endogeneity, that is, the existence of unobservable individual characteristics, which jointly determine both income and health, as for example, the social background of the individual; 2. Reverse causality which implies that the direction of the causation could be happening from health to SE status or vice versa; 3. Identification of the most suitable measures of income and health that should be used in the analysis, which is usually an issue related to data constraints for some surveys.

Health dynamics have been recently studied by Contoyannis, Jones and Rice (2004), work that we use as our benchmark for our analysis. They base their study on the British Household Panel Survey (BHPS), focusing on three issues. Firstly, they check the relative contribution of state dependence and heterogeneity in explaining the dynamics of health. Secondly, they analyse whether there is evidence of health-related attrition in the sample and its consequences. Finally, they explore the relationship between SE characteristics (education and income) and a measure of self-reported health (SAH) included in the BHPS.

Contoyannis, Jones and Rice (2004) model SAH by using dynamic Pooled Ordered Probit models and specifying the functional form of a latent variable representing "true health". They deal with correlated individual effects and the initial conditions problem by focusing on Wooldridge's (2005) approach, while they perform a variable-addition test for attrition bias (Verbeek & Nijman, 1992) and apply inverse probability weighting (IPW) to adjust for attrition when they estimate pooled models.

For the purpose of their study, they compare the estimates obtained when using a balanced panel (BP) and an unbalanced panel (UP). Besides, all available data is used to evaluate the impact of attrition. To analyse non-response and attrition bias, they compare the number of observations across waves together with the bivariate relationship between attrition rates and SE characteristics.

They found evidence of attrition related to health in the data, but estimates of state dependence and of the SE gradient in health has not been distorted by this bias. Heterogeneity was found to account for around 30% and positive state dependence was shown in the results of the models.

A more detailed study on attrition bias has been presented by Jones, Koolman and Rice (2006). They analyse health-related non-response using a categorical variable of SAH, as it may have consequences in dynamic models. They use a BP and UP and correct for non-response using IPW (Wooldridge, 2002).

The objective of Jones, Koolman and Rice (2006) is to find evidence of health-related non-response in panel data and the consequences it has for modelling the relationship between SE status and SAH. For that purpose, they describe the pattern of non-response due to health that is revealed by the BHPS and the ECHP dataset.

They use 9 waves available of the BHPS and they concentrate in SAH, self-reported functional limitations, specified health problems and an indicator of being registered as a disable person. From the ECHP, they use and indicator of any limitation and severe limitation.

They found evidence of non-response related to health in both datasets, but with a limited impact on the estimates of health dynamics and estimates of the relationship between SE characteristics and SAH.

Evidence on income-related inequalities in health problems for the 8 periods covered by the ECHP (1994 – 2001) have been provided at the EU-15 level (Hernández-Quevedo et al, 2006). In particular, income-related inequalities in health limitations have been found in the 14 Member States considered, in both the short and the long run. These inequalities favour the richest individuals in each society and they show an increasing pattern in most of the countries. This evidence justifies measures of public policies to reduce these inequalities in the EU[1]. Besides, this study suggests the use of a longitudinal perspective when measuring and interpreting inequalities in health, against the approach usually found in the literature that focus on cross-sectional data.

However, this study presents some limitations, as only health inequalities due to differences in income levels are studied, without taking into account other socioeconomic factors as education and job status.

Our aim is to extend the above study, by quantifying the causal effect of SE characteristics in two indicators of health limitations.

[1] Action to reduce health inequalities in EU aims: 1. To improve everyone's level of health closer to that of the most advantaged; 2. To ensure that the health needs of the most disadvantaged are fully addressed; 3. to help the health of people in countries and regions with lower levels of health to improve faster. *European Commission.*

2.1. New contributions

In our study, we work with a panel dataset that contains 15 European countries and 8 waves of data. The three issues raised by Contoyannis et al (2004) are analysed: the relative contribution of state dependence and heterogeneity in explaining dynamics of health, the initial conditions problem and the existence and consequences of health-related attrition. The latest issue has been already study by Contoyannis et al (2004), using the BHPS. They argue that "failing to account for attrition leads to misleading estimates of health dynamics and of the relationship between health and SE characteristics". They found evidence of health-related attrition in their dataset but this did not distort the estimates of the relationship between SAH and SE status, using the BHPS. The same argument was used by Jones, Koolman and Rice (2006) to study attrition bias, finding the same results using the ECHP dataset.

In our study, we add job status as one of the SE characteristics that can contribute to report health limitations in the daily activity. This constitutes the main difference with other studies that have focus on income and education, analysing a categorical measure of self-reported health.

3. The ECHP-UDB data

The *European Community Household Panel Users Database* (ECHP-UDB) is a standardised annual longitudinal survey, designed and coordinated by the European Commission's Statistical Office (EUROSTAT). It provides 8 waves (1994 - 2001) of comparable micro-data about living conditions in the European Union Member States (EU-15). The survey is based on a standardised questionnaire that involves annual interviewing of individuals aged 16 and older from a representative panel of households (Peracchi, 2002). National Data Collection Units implemented the survey in each of the member countries. Approximately, 60,000 households and 130,000 adults were interviewed at each wave. The survey covers a wide range of topics including demographics, income, social transfers, individual health, housing, education and employment. The information provided in the ECHP-UDB can be compared across countries and over time, making it an attractive dataset for the purpose of our study.

The first wave included all EU-15 Member States with the exception of Austria and Finland. Austria joined in 1995 and Finland, in 1996. For the first three waves, the ECHP ran parallel to existing national panel surveys in Germany, Luxembourg and the United Kingdom. From the fourth wave onwards, the ECHP samples were replaced by data harmonized ex-post from these three surveys. Hence, there were two versions of the ECHP database for Germany, Luxembourg and the United Kingdom. Although Sweden did not take part in the ECHP, the Living Conditions Survey[2] is included in the UDB, together with comparable versions of the British Household Panel Survey (BHPS), the German Socioeconomic Panel (GSOEP) and the Panel Survey for Luxembourg (PSELL)[3].

For the purpose of this study, we have included in our analysis those Member States of the EU contained in the ECHP that have 8 waves of available data. These are: Denmark, The Netherlands, Belgium, France, Ireland, Italy, Greece, Spain and Portugal.

[2] Note however that the data for Sweden is not longitudinal, and has been derived from repeated cross-sections. We do not use data for Sweden.

[3] Data for Germany, Luxembourg and United Kingdom are taken from the original ECHP survey.

Sample and variables

We need a full set of waves for each individual and we use a balanced sample of respondents, which implies that only individuals from the first wave who were interviewed in each subsequent wave are included in the analysis[4]. Table 1 shows the sample size for each country, for the whole sample and split by gender. For most countries, the sample size is between 20,000 and 50,000 observations. Exceptions are Spain and Italy with both having notably larger samples and Luxembourg and the United Kingdom with notably smaller samples.

[Insert Table 1 around here]

Health limitations

The ECHP-UDB dataset contains some limited information on health outcomes and health care utilisation. We use the information on health limitations, in particular responses provided to the question[5]: "Are you hampered in your daily activities by any physical or mental health problem, illness or disability?". Three possible answers are available for the respondent: "Yes, severely", "Yes, to some extent" and "No". In the ECHP-UDB, this information is provided for all countries and waves that we consider for our analysis[6]. We focus on two binary measures of health problems that have been derived from the responses to the health limitations question. From these responses, two dummy variables are constructed. The first variable labelled HAMP1, represents an indicator of any limitations (severe or to some extent) versus no limitations; the second dummy (HAMP2) represents an indicator of severe limitations versus no limitations or limited to some extent.

Explanatory variables

Five variables represent marital status (*Widowed*, *Single*, *Divorced*, *Separated*) with *Married* as the reference category. Three dummy variables have been constructed to represent maximum level of education attained: *Tertiary* (Third level), *Secondary* (second stage of secondary level) and *Primary* (less than second stage of secondary education), with *Tertiary* being the base case for the education variables. The size of the household (*HHSize*) and the number of children in the household by age (*nch04*, *nch511*, *nch1218*), are also included in the analysis. The income variable is the logarithm of equivalised real income, adjusted using the Purchasing Power Parities and the Consumer Price Index. It is equivalised by the OECD-modified scale to adjust for household size and composition. There are six possible categories for job status: *Self-employed*, *Unemployed*, *Retired*, *Housework* and *Inactive*, with *Employed* individuals being the reference case. Individuals have been grouped by age and sex, with a man (woman) aged between 16 and 25 being the

[4] Care should be taken when interpreting the results as the respondents in the balanced panel may not be representative of the full sample. Jones, Koolman and Rice (2005) have provided evidence of health-related non-response in the ECHP but they also find that estimates of the association between health and socioeconomic status are robust with or without adjustments for non-response.
[5] The question is coded PH003A in the ECHP-UDB.
[6] Although the question was asked similarly in all the countries where the data was available, the French case is an exception as the question was reworded for the full panel (1994 – 2001) from "… hampered by any chronic, physical or mental health problem, illness or disability?" to "Gêné par une maladie chronique, un handicap?".

reference case. A vector of time dummies is also included in the analysis. See Table 2 for a full list of variables used in this study.

[Insert Table 2 around here]

Descriptive Analysis using a Balanced Panel

Figure 1 shows the distribution of HAMP for all countries and it shows that health limitations present a similar distribution across countries, with most individuals reporting not perceiving any or severe limitations in their daily activity.

[Insert Figure 1 around here]

Figures 2 and 3 show the distribution of HAMP1 and HAMP2 respectively, for each country. For the variable HAMP1, the country with the highest percentage of individuals who report any limitation is Finland at 28.2%, followed by Portugal (25.6%) and the United Kingdom (25.2%). The country with the lowest percentage is Italy (12.6%), followed by Belgium (14.8%) and Ireland (16.2%). Similar results are found for the variable HAMP2. Portugal has the highest percentage of individuals who report being severely hampered (10.3%), followed by France (9.5%) and Finland (7.6%), while Ireland, Italy and Belgium have the lowest percentages at 3.4%, 4.3% and 4.6%, respectively.

[Insert Figure 2 around here]

[Insert Figure 3 around here]

Table 3 shows the percentage of individuals who report either any or severe limitations across income quintiles. Minimum and maximum percentages are highlighted. These range from 6.3% of respondents who report some health limitations in the fifth income quintile in Italy to 26% in the first income quintile in the United Kingdom. The range for severe health limitations goes from 1.4% in the fifth quintile for Ireland to 15.4% in the second income quintile in Portugal.

Country-specific results show a clear association between income and health. In general, there exists a gradient across income quintiles in the reporting of both severe and any health limitations, such that a higher proportion of respondents in lower income quintiles report limitations compared to respondents from higher quintiles. Further, there is variation across countries in the observed income gradients. For example, for Portugal the gradient ranges from 15.4% of respondents reporting severe limitations in the second quintile to 5.5% in the fifth quintile. For Italy, the range is 5.2% in the first quintile to 2.7% in the fifth quintile. Similarly there is variation across income quintiles in the proportion of respondents reporting health limitations to some extent. For Luxembourg, the proportion ranges from 20.7% in the lowest quintile to 11.5% in the highest. This is in contrast to Italy where the corresponding figures are much lower at 9.2% and 6.3%, respectively.

[Insert Table 3 around here]

From Table 4, it can be seen that there is a common pattern for all the countries listed in the table. The highest percentage of individuals who report limitations in daily activity to

some extend are either employed or retired for several countries (Germany, Denmark, Belgium, Luxembourg, France, United Kingdom, Portugal, Austria and Finland). For The Netherlands, Ireland, Italy, Greece and Spain, the highest percentage correspond to those who are either retired, doing housework or inactive.
For the case of those reporting severe limitations in daily activity, the highest percentage correspond to those who are retired excepting The Netherlands, where employed and housework concentrate the highest percentages. For Italy and Spain, the percentage of those inactive and reporting severe limitations is large (31.18% and 34.23%, respectively). For Italy, those whose job status is housework and report severe limitations is also significative (22.91%).

[Insert Table 4 around here]

1. Methods

The general latent variable specification for a binary choice model in a dynamic context is given by expression (1):

$$h_{it}^{*} = x_{it}^{\prime}\beta + \gamma h_{it-1} + \eta_i + \varepsilon_{it} \qquad (1)$$

Where x_{it} is the set of explanatory variables, η_i is a time-invariant individual effect and ε_{it} is a time and individual-specific error term.

Hence,

$$\begin{aligned} h_{it} &= 1, \text{ if } h_{it}^{*} > 0 \\ h_{it} &= 0, \text{ otherwise} \end{aligned} \qquad (2)$$

If we assume that the distribution of ε_{it} is symmetric with distribution function F(.),

$$P(h_{it}=1 \mid x_{it}, \beta, \gamma, h_{it-1}, \eta_i) = P(\varepsilon_{it} > -x'_{it}\beta - \gamma h_{it-1} - \eta_i) = F(x'_{it}\beta + \gamma h_{it-1} + \eta_i) \qquad (3)$$

In our study, we assume a standard normal distribution, that is, a probit model. Hence,

$$\Pr[h_{it}=1 \mid x_{it}, \beta, \gamma, h_{it-1}, \eta_i] = \Phi(x_{it}^{\prime}\beta + \gamma h_{it-1} + \eta_i)\,, \qquad (4)$$

where $\Phi(.)$ is the standard normal cumulative distribution function (cdf).
If conditional independence is assumed, the joint density for the i^{th} observation $h_i = (h_{i1}, \ldots, h_{iT})$ is:

$$f(h_i \mid x_i, \beta, \gamma, h_{it-1}, \eta_i,) = \prod_{t=1}^{T} F(\eta_i + x_{it}^{\prime}\beta + \gamma h_{it-1})^{h_{it}} (1 - F(\eta_i + x_{it}^{\prime}\beta + \gamma h_{it-1}))^{1-h_{it}} \qquad (5)$$

Several models have been used in empirical studies of the relationship between health and socioeconomic characteristics. In this paper, we consider different estimation models: pooled probit and logit models, random effects (RE) probit and logit models, together with complementary log-log and RE complementary log-log models.

4.1. Pooled Specification

The pooled specification does not take into account that the panel dataset contains repeated observations, that is, it pools all the observations together, without considering that individuals are measured repeatedly, from wave 1 to wave 8.
The log-likelihood used for the pooled model assumes that the observations are independent across waves and simply uses the product of their marginal distributions. Most likely, observations are correlated within individuals, which implies that the joint distribution is misspecified and hence, estimates are not Maximum Likelihood estimators, but Quasi-Maximum Likelihood estimators (QMLE) (Wooldridge, 2002).
The estimates resulting from the QMLE are consistent, although it does not take into account the structure of the error term.

4.2. Random Effects Specification

The Random Effects (RE) specification assumes that both components of the error term (η_i, ε_{it}) are normally distributed and that both are independent of the x's, which is a strong assumption. Different approaches have been shown in the literature trying to relax this assumption.
Mundlak (1978) suggested an approach to relax the assumption of independency, which specifies this relationship as a linear regression of the mean value of the explanatory variables, that are averaged over t for a given i:

$$\eta_i = \bar{x}_i'\pi + \xi_i \tag{6}$$

where ξ_i is independent and identically distributed. The vector π will equal 0 if and only if the explanatory variables are uncorrelated with the effects.
Chamberlain (1984) presents a correlated random effects model to deal with individual RE that are correlated with the explanatory variables. This approach consists in specifying this relationship as a linear regression of the value of the explanatory variables in all the waves of the panel. If there is sufficient within-individual variation, it is possible to obtain separate estimates of the β's and to disentangle the correlation between the x's and the time-invariant individual-specific effect η_i.
Chamberlain's model specifies:

$$\eta_i = x_{1i}'\pi_1 + \ldots + x_{Ti}'\pi_T + \xi_i\,, \tag{7}$$

where $\xi_i | x_i \sim N(0, \sigma_\eta^2)$.

Wooldridge (2005) provides an approach to deal with correlated individual effects and the initial conditions problem in dynamic, nonlinear unobserved effects probit model. It consists in finding the distribution conditional on the initial value.
There are two factors that can be problematic: 1. the starting point of a survey is not the beginning of a process, 2. individuals inherit different unobserved and time-invariant characteristics which affect outcomes in every period, issue that can lead to endogeneity bias in dynamic models with covariance structures that are not diagonal (Contoyannis et al, 2004). The approach provided by Wooldridge (2005) models the distribution of the unobserved effect conditional on the initial value and any strictly exogenous explanatory variables. The resulting likelihood function is based on the joint distribution of the observations conditional on the initial observations (conditional maximum likelihood).

4.3. Complementary log-log binary choice model

The logit model is one of the most commonly used binary outcome models. In this case, F(.) is the cdf of the logistic distribution. The less used complementary log-log model arises if F(.) is the cdf of the extreme value distribution. It is frequently used when the probability of an event is very small or very large. Unlike logit or probit, the complementary log-log function is asymmetrical.

The probability of an individual i reporting health limitations is given by expression (8):

$\Pr[h_{it}=1 \mid x] = C(x'\beta) = 1-\exp(-\exp(x'\beta))$ (8)

The marginal effect can be written as :

$(dp/dx_j) = \exp(-\exp(x'\beta)) \exp p(x'\beta)\beta_j$ (9)

The log-likelihood function for complementary log-log is given by expression (10):

$$\ln L = \sum w_j \ln F(x_j b) + \sum w_j \ln\left(1 - F\left(x_j b\right)\right) \quad (10)$$

where F(z) = 1-exp(-exp(z)). w_j denotes optional weights.

2. Estimation strategy

We explore the relationship between health and SE status, in particular, the relationship between reporting health limitations in the daily activity and household income, education and job status, among other SE controls as marital status, household size, number of children in the household within a certain age interval and age-sex groups.

We use dynamic panel specifications on a balanced sample to model HAMP1. Previous report on health limitations is included in our specification to capture state dependence and reduce bias due to reverse causality (see Adams et al, 2003). Hence, this model can be seen as a first-order Markov process. These models can be regarded as reduced form specification, that is, variables such as medical care and lifestyle are not included (Contoyannis et al, 2004).

Our model presents the following specification for its binary latent variable:

$h_{it}^{*} = x_{it}'\beta + \gamma h_{it-1} + \delta h_{io} + u_{it}$ (11)

Where u_{it} is a two-component error term, made of a time-invariant individual effect (η_i) plus a time and individual-specific error term (ε_{it}), where ε_{it} follows a Normal (0, 1) and is independent of the x's and η_i. Hence,

$u_{it} = \eta_i + \varepsilon_{it}$ (12)

We apply Wooldridge (2005) approach to deal with the initial conditions problem by including the initial value of health limitations h_{io}.

To allow for the possibility that the observed regressors may be correlated with the individual effect we parameterize the individual effect (Mundlak, 1978; Chamberlain, 1984; Wooldridge, 2005).

We implement Wooldridge (2005) approach by parameterizing the distribution of the individual effects as follows:

$\eta_i = \pi x_{io} + \delta h_{io} + \gamma_i$ (13)

We substitute the expression for η_i (13) in expression (11) and we obtain the model to be estimated (14):

$h_{it} = \alpha h_{it-1} + \beta x_{it} + \delta h_{io} + \pi x_{io} + \gamma_i + \varepsilon_{it}$, (14)

where h_{it-1} is the lagged dependent variable and x_{it} includes the lagged values of the regressors and the time-invariant variables. α, β, δ and π are parameters to be estimated.

We include: current values of education, household size, number of children by age and age groups, initial values of log of income and job status, and lagged values of marital status, log of income and job status.
We check for the reliability of the approximation of points in the quadrature provided for the random effects (probit, logit and complementary log-log) specification that equals 12 by default. We find significant relative differences in the estimates when the number of quadrature points is increased from 8 to 16, through 12. Hence, we use 24 points for our estimations, improving the accuracy of the estimates.

Information Criteria

To decide which model best suits the data, we use 2 criteria: Akaike Information Criterion and Bayesian Information Criterion. The information criteria are log-likelihood criteria with degrees of freedom adjustment. They capture the trade-off between the model fit (measured by maximised log-likelihood value) and the principle of parsimony that favours a simple model, penalising model complexity. AIC imposes a higher penalty than BIC.

$AIC = -2LnL + 2q$

$BIC = -2LnL + (LnN)q$

where q represents number of coefficients and N denotes number of observations.

Marginal Effects

The model with the minimum AIC and BIC is considered to obtain the marginal effects for further interpretation.

3. Results

If we compare the results in Table 5 & Table 6 for both indicators of health limitations, we can see that the pooled probit specification is the one that has the minimum value of both information criteria for all the countries considered. Hence, we calculate the marginal effects for this specification for both hamp1 and hamp2.
Table 7 and Table 8 show the results for the marginal effects for the pooled probit specification for hamp1 and hamp2, respectively.

For both indicators of health limitations, it is possible to see that the signs are negative as expected for highest level of education achieved and household income. There are some exceptions, although these estimates are not statistically significant at a 5% or 10% significance level. Besides, it can be seen that the estimates for the variable indicating if an individual is self-employed does not follow a specific trend. The coefficient corresponding to the lagged value of health limitations is statistically significant for all the cases and it is significantly large, indicating a high persistence of health limitations.
For hamp1, the largest coefficients for health limitations in t-1 correspond to Portugal and The Netherlands, while for Italy and Spain the values are the lowest among the countries considered. For hamp2, the greatest estimates for lagged health limitations correspond to France and Portugal, while Ireland and Spain present the lowest values.
If the marginal effects for hamp1 are compared with those corresponding to hamp2, it is possible to see that hamp1 presents higher marginal estimates than hamp2, hence, showing greater persistence if the individual has any limitation than in the case that he reports severe limitations.

4. Conclusions

We present a dynamic approach taking into account the 8 waves available of the European Community Household Panel Users Database (ECHP-UDB). Differing from other studies on this issue, we focus on two indicators of health limitations (any limitation and severe limitation) and we include job status as explanatory variables in our analysis.
The provisional conclusions are that the probit model, in particular, the pooled probit model is the adequate specification for our sample, according to the Information Criteria that have been calculated here. From the marginal effects, we have seen that including state dependence is important, as we show that the estimated coefficients on lagged health limitations are large and highly statistically significant.

References

Adams, P, MD Hurd, D McFadden, A Merril, T Ribeiro (2003). Healthy, wealthy, and wise. Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112: 3 - 56

Benzeval, M and K Judge (2001). Income and health: the time dimension. *Social Science & Medicine*, 52: 1371 - 1390

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 (3): 375 - 399

Blundell, R and I Preston (1995). Income, expenditure and the living standards of UK households. *Fiscal Studies*, 16 (3): 40 - 54

Chamberlain, G (1984). Panel data In *Handbook of Econometrics*, Vol. 1, Griliches, Z, MD Intriligator (eds.). North Holland: Amsterdam, 1247 - 1318

Contoyannis, P, AM Jones and N Rice (2004). The dynamics of health in the British Household Panel survey. *Journal of Applied Econometrics*, 19: 473 - 503

Deaton, AS and CH Paxon (1998). Aging and inequality in income and health. *The American Economic Review*, 88 (2): 248 – 253

Ettner, SL (1996). New evidence on the relationship between income and health. *Journal of Health Economics*, 15: 67 - 85

Fritjers, P, JP Haisken-DeNew, MA Shields (2003). Estimating the causal effect of income on health: evidence from post reunification East Germany. *Discussion Papers No. 465, Centre for Economic Policy Research*

Jones, AM, X Koolman, N Rice (2006). Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. *Journal of the Royal Statistical Society*, *Series A*, forthcoming

Mundlack, Y (1978). On the pooling of time series and cross section data. *Econometrica*, Vol. 46, No. 1: 69 - 85

Peracchi, F (2002). The European Community Hosehold Panel: A Review, *Empirical Economics*, 27: 63 - 90

Smith, JP (2004). Unraveling the SES health connection. *Working Paper 04-02 The Institute for Fiscal Studies*

Verbeek, M and TE Nijman (1992). Testing for selectivity bias in panel data models. *International Economic Review* 33: 681 - 703

Williams, A., R Cookson (2000). Equity in Health. In Aj Culyer and JP Newhouse (Eds.). *Handbook of Health Economics*. Volume 1b., Elsevier.

Wooldridge, JM (2002). Inverse probability weighted M-estimators for sample stratification, attrition and stratification. *Portuguese Economic Journal*, 1: 117 - 139

Wooldridge, JM (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20: 39 - 54

Table 1: Sample size for each country considered in the analysis

Wave	D	DK	NL	B	L	F	UK	Irl	I	EL	E	P	A	Fin
1	8,036	2,536	4,656	3,008	1,779	7,226	5,382	2,748	9,539	6,384	7,549	7,348	-	-
2	8,036	2,536	4,656	3,008	1,779	7,226	5,382	2,748	9,539	6,384	7,549	7,348	4,001	-
3	8,036	2,536	4,656	3,008	1,779	7,226	5,382	2,748	9,539	6,384	7,549	7,348	4,001	3,893
4	-	2,536	4,656	3,008	-	7,226	-	2,748	9,539	6,384	7,549	7,348	4,001	3,893
5	-	2,536	4,656	3,008	-	7,226	-	2,748	9,539	6,384	7,549	7,348	4,001	3,893
6	-	2,536	4,656	3,008	-	7,226	-	2,748	9,539	6,384	7,549	7,348	4,001	3,893
7	-	2,536	4,656	3,008	-	7,226	-	2,748	9,539	6,384	7,549	7,348	4,001	3,893
8	-	2,536	4,656	3,008	-	7,226	-	2,748	9,539	6,384	7,549	7,348	4,001	3,893
Men	11,640	9,776	16,928	10,808	2,571	26,936	7,119	10,512	36,840	23,224	27,712	26,960	13,370	11,484
Women	12,468	10,512	20,320	13,256	2,766	30,872	9,027	11,472	39,472	27,848	32,680	31,824	14,637	11,874
Total	24,108	20,288	37,248	24,064	5,337	57,808	16,146	21,984	76,312	51,072	60,392	58,784	28,007	23,358

Table 2: Explained and explanatory variables

HAMP	Hampered in daily activities by any physical or mental health problem, illness or disability:
	1 if severely hampered by any health problem
	2 if some extend hampered by any health problem
	3 if not hampered by any health problem
HAMP1	1 if severely hampered or to some extend by any health problem, 0 otherwise
HAMP2	1 if severely hampered by any health problem, 0 otherwise
SEPARATED	1 if separated, 0 otherwise
DIVORCED	1 if divorced, 0 otherwise
WIDOWED	1 if widowed, 0 otherwise
NVRMAR	1 if never married, 0 otherwise
SECONDARY	1 if third level, 0 otherwise
PRIMARY	1 if less than second stage of secondary level, 0 otherwise
HH_SIZE	Number of people in household including respondent
NCH04	Number of children in household aged 0 - 4
NCH511	Number of children in household aged 5 - 11
NCH1218	Number of children in household aged 12 - 18
INCOME	Equivalised total net household income (PPP & CPI)
AGE2635M	Men with age between 26 and 35
AGE3645M	Men with age between 36 and 45
AGE4655M	Men with age between 46 and 55
AGE5665M	Men with age between 56 and 65
AGE6675M	Men with age between 66 and 75
AGE7685M	Men with age between 76 and 85
AGE86M	Men with 86 years old or more
AGE1625F	Women with age between 16 and 25
AGE2635F	Women with age between 26 and 35
AGE3645F	Women with age between 36 and 45
AGE4655F	Women with age between 46 and 55
AGE5665F	Women with age between 56 and 65
AGE6675F	Women with age between 66 and 75
AGE7685F	Women with age between 76 and 85
AGE86F	Women with 86 years old or more
SELFEMPLOY	1 if self-employed, 0 otherwise
UNEMPLOYED	1 if unemployed, 0 otherwise
RETIRED	1 if retired, 0 otherwise
HOUSEWORK	1 if doing housework, looking after children or other persons, 0 otherwise
INACTIVE	1 if other economically inactive, 0 otherwise

Table 3: Percentage of health limitations by income quintiles

Country	Limitations to some extent Income quintiles					Severe limitations Income quintiles				
	1	2	3	4	5	1	2	3	4	5
Germany	17.72	17.15	15.67	14.33	14.90	9.63	7.34	5.72	5.53	4.72
Denmark	20.38	17.54	16.64	13.81	11.23	10.75	7.12	3.43	2.67	2.08
Netherlands	18.61	17.36	15.53	14.86	13.66	10.25	9.08	6.73	5.55	5.09
Belgium	14.46	10.70	8.94	8.79	8.71	9.53	5.57	3.17	2.18	2.55
Luxembourg	20.65	18.36	18.91	12.91	11.53	7.14	5.16	4.58	4.16	2.34
France	16.69	15.01	12.98	10.18	10.18	14.11	11.90	10.52	5.65	5.65
UK	25.76	21.77	17.10	13.66	14.54	9.76	10.48	7.06	4.22	2.13
Ireland	17.24	20.35	13.09	10.62	7.98	6.82	6.26	3.26	1.85	1.44
Italy	9.18	9.94	9.09	7.91	6.26	5.22	5.22	4.98	3.97	2.72
Greece	14.39	11.81	9.73	9.50	6.43	12.26	9.44	7.55	6.51	3.49
Spain	14.71	15.51	13.44	10.49	7.01	7.23	7.36	7.08	5.38	2.59
Portugal	19.35	18.53	15.93	14.25	11.14	14.30	15.36	11.34	8.43	5.50
Austria	18.30	14.44	12.23	11.40	11.25	8.18	5.33	4.97	3.90	3.32
Finland	22.21	22.09	19.87	20.46	18.63	10.09	9.78	6.80	6.38	5.61

Note: Both the highest and lowest percentages of responses by income quintiles across countries have been highlighted in this table

Table 4: Percentage of health limitations by job status

Country	Limitations to some extent Job status						Severe limitations Job status					
	E	S	U	R	H	I	E	S	U	R	H	I
Germany	36.82	2.88	5.49	39.51	12.63	0.73	15.95	1.67	6.21	63.42	8.55	3.46
Denmark	44.81	3.60	5.20	39.08	2.40	1.48	13.11	2.29	3.30	74.43	1.47	4.77
Netherlands	35.47	2.88	8.48	2.94	36.03	11.88	23.45	1.61	11.76	2.81	40.10	18.77
Belgium	29.24	3.89	6.57	40.31	14.09	4.58	12.98	2.47	7.95	49.82	13.15	13.25
Luxembourg	32.38	3.90	1.72	36.39	22.27	0.80	20.73	2.85	5.28	45.93	19.11	5.69
France	31.85	3.79	4.57	45.38	12.51	0.35	17.84	2.82	4.61	53.92	19.46	0.31
UK	25.71	5.96	4.09	38.75	19.01	4.22	7.72	2.57	1.98	55.79	15.64	13.35
Ireland	15.90	9.27	4.58	21.33	35.68	11.89	7.80	4.57	2.55	27.15	26.08	31.18
Italy	18.39	8.23	3.75	41.79	22.44	4.08	7.97	4.22	2.57	49.22	22.91	12.68
Greece	8.76	14.45	2.68	46.80	25.56	1.67	3.19	6.80	1.65	58.50	20.99	8.77
Spain	9.30	6.16	3.81	29.19	35.03	15.95	4.76	2.63	2.34	31.19	24.70	34.23
Portugal	21.27	15.79	3.30	38.02	15.25	5.78	9.36	7.88	2.44	52.12	11.45	16.40
Austria	20.88	8.67	3.61	47.35	18.35	0.53	8.91	4.70	3.39	59.30	20.87	2.76
Finland	35.27	12.54	6.60	41.35	1.10	0.27	15.79	10.11	4.16	66.52	0.79	1.40

Note: 6 categories have been used to reflect job status: employed (E), self-employed (S), unemployed (U), retired (R), housework (H) and inactive (I).

Figure 1: **Distribution of health limitations (HAMP) for each country**

Figure 2: **Percentage of individuals hampered (HAMP1), across the Member States**

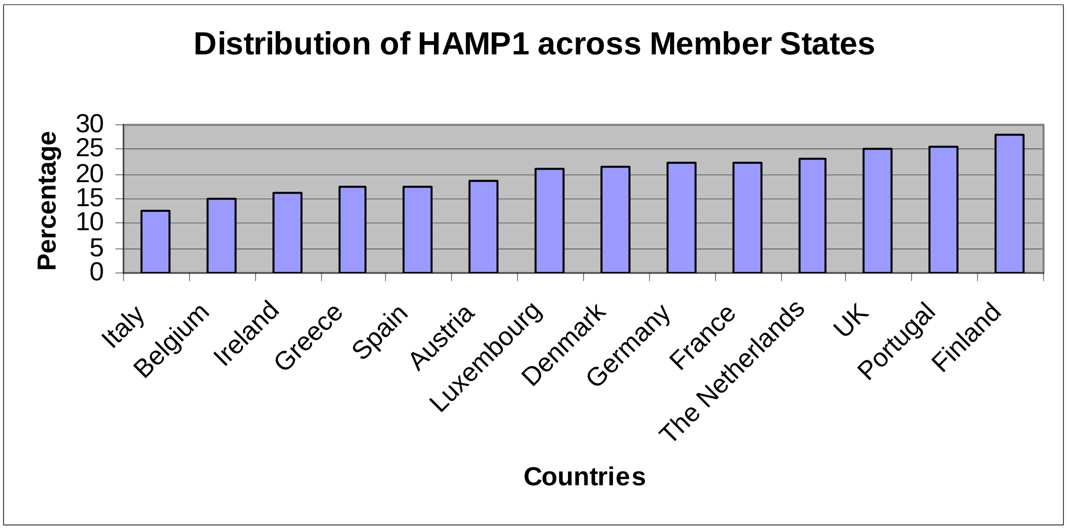

Figure 3: Percentage of individuals severely hampered (HAMP2) across the Member States

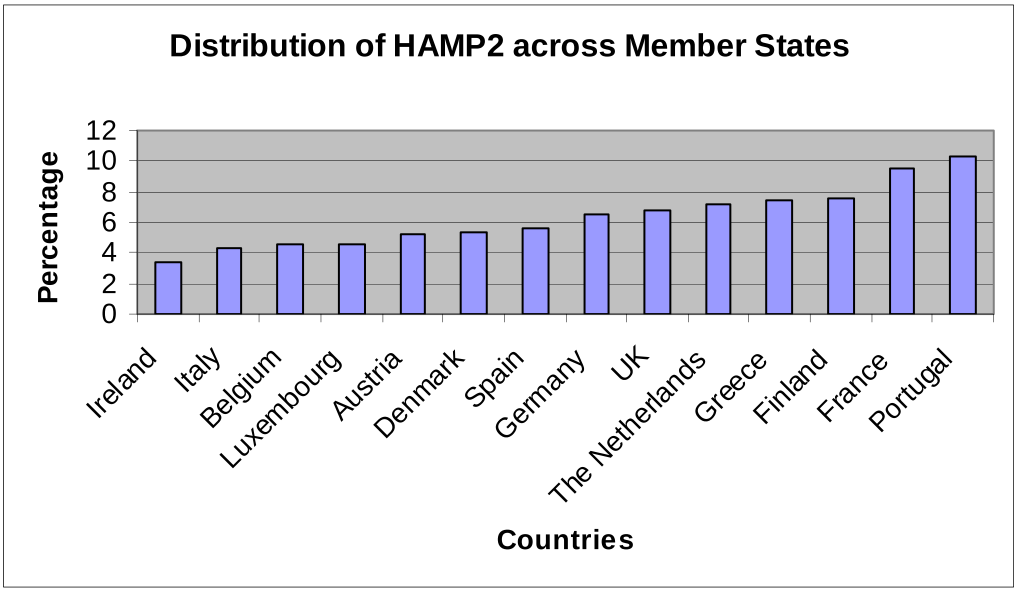

Table 5: AIC & BIC results for different specifications using HAMP1

		AIC	BIC
DK	PP	11971.16	12332.17
	REP	13789.44	14106.09
	PL	11979.56	12337.57
	REL	13810.34	14127
	CLL	12005.84	12363.85
	RE-CLL	13795.98	14112.63
NL	PP	21826.28	22202.81
	REP	26594.9	26926.78
	PL	22645.85	23022.37
	REL	26602.22	26934.1
	CLL	22680.03	23056.56
	RE-CLL	26589.76	26921.64
B	PP	10453.37	10818.8
	REP	12574.83	12898.11
	PL	10474.87	10840.3
	REL	12603.77	12927.05
	CLL	10579.77	10945.2
	RE- CLL	12648.83	12972.11
F	PP	32187.85	32581.67
	REP	39061.68	39419.66
	PL	32209.79	32603.61
	REL	39029,49	39387,46
	CLL	32456.53	32850.35
	RE-CLL	39008,29	39366,26

Table 5: AIC & BIC results for different specifications using HAMP1 (cont.)

		AIC	BIC
Irl	PP	10681.61	11043.23
	REP	12550,32	12870.18
	PL	10714,17	11075.79
	REL	12558,54	12878.4
	CLL	10816.73	11178.35
	RE-CLL	12676,75	12996.61
I	PP	28445.79	28863.29
	REP	36798.27	37167.48
	PL	28474.61	28892.11
	REL	36845.84	37215.05
	CLL	28690.34	29107.84
	RE-CLL	37243.38	37612.58
EL	PP	26826.2	27226.25
	REP	31675.98	32029.38
	PL	26899.08	27299.14
	REL	31690.39	32043.79
	CLL	27070	27470.05
	RE-CLL	31827.61	32181.01
E	PP	31665.4	32073.21
	REP	36317.61	36677.78
	PL	31825.58	32233.38
	REL	36361.34	36721.51
	CLL	32270.23	32678.04
	RE-CLL	36648.86	37009.03
P	PP	34644.8	35050.55
	REP	42930.77	43289.66
	PL	34705,74	35111,48
	REL	42954.05	43312.93
	CLL	35085.77	35491.51
	RE-CLL	43233.7	43592.58

Table 6: AIC & BIC results for different specifications using HAMP2

		AIC	BIC
DK	PP	4687.831	5045.839
	REP	5112.71	5429.365
	PL	4715.315	5073.323
	REL	5115.428	5432.082
	CLL	4776.912	5134.92
	RE-CLL	5127.975	5444.629
NL	PP	11826.28	12202.81
	REP	13219.61	13551.49
	PL	11862.34	12238.87
	REL	13233.97	13565.85
	CLL	11953.23	12329.76
	RE-CLL	13234.2	13566.08

Table 6: AIC & BIC results for different specifications using HAMP2 (cont.)

		AIC	BIC
B	PP	4941.878	5298.021
	REP	5656.982	5980.262
	PL	4987.776	5343.919
	REL	5674.86	5998.14
	CLL	5049.503	5405.646
	RE-CLL	5688.452	6011.733
F	PP	19136.92	19530.74
	REP	22880.37	23238.34
	PL	19212.04	19605.86
	REL	22912,79	23270,76
	CLL	19412.01	19805.83
	RE-CLL	22933.66	23291.63
Irl	PP	3810.951	4172.574
	REP	4305.038	4624.893
	PL	3834,336	4195.959
	REL	4305,096	4624.952
	CLL	3867,635	4229.258
	RE-CLL	4323,865	4643.72
I	PP	13591.67	14009.17
	REP	16774.13	17143.33
	PL	13726.9	14144.4
	REL	16828.58	17197.79
	CLL	13892.29	14309.79
	RE-CLL	16918.44	17287.64
EL	PP	16351.66	16751.72
	REP	18691.69	19045.09
	PL	16411.6	16811.65
	REL	18675.31	19028.71
	CLL	16498.67	16898.72
	RE-CLL	18708.91	19062.31
E	PP	16730.75	17138.56
	REP	18424.54	18784.71
	PL	16852.35	17260.16
	REL	18464.7	18824.86
	CLL	16996.33	17404.13
	RE-CLL	18576.87	18937.04
P	PP	22234.27	22640.01
	REP	25614.87	25973.75
	PL	22355,58	22761,32
	REL	25618.5	25977.38
	CLL	22582.71	22988.46
	RE-CLL	25682.82	26041.7

Table 7: Marginal Effects, Pooled Probit model, hamp1

	DK	NL	B	F	Irl	I	EL	E	P
hamp1_lag	0.466*	.471*	.399*	.451*	.412*	0.365*	.394*	.258*	.506*
primary	-0.03*	-.050*	-.026*	-.047*	-.012	-.003	-.020*	-.022*	-.010
secondary	-0.008	-.025*	-.009	-.020*	-.004	-.005**	-.007	-.006	.002
ln_inc_lag	-0.004	-.015*	.004	-.016*	-.016*	-.00005	-.006*	-.011*	-.026*
selfemploy_lag	0.0001	-.031**	-.007	-.006	-.008	.0005	-.004	-.013	.010
unemployed_lag	0.046*	.033	.036*	.040*	.066*	.006	.036*	.042*	.053*
retired_lag	0.1*	.014**	.022*	.037*	.019	.014*	.042*	.057*	.089*
housework_lag	-0.02	.016**	.035*	.067*	-.010	.005	.027*	.048*	.056*
inactive_lag	0.131*	.044*	.183*	.003	.261*	.079*	.173*	.159*	.131*

Table 8: Marginal Effects, Pooled Probit model, hamp2

	DK	NL	B	F	Irl	I	EL	E	P
hamp2_lag	.276*	.334*	.227*	.344*	.209*	0.242*	.257*	.122*	.360*
primary	-.011*	-.014*	-.011*	-.016*	-.004**	-.002*	-.009*	-.007*	-.001
secondary	-.003	-.003	.001	-.006**	-.002	-.002*	-.004**	-.001	.00001
ln_inc_lag	-.004**	-.007*	.0001	-.005*	-.002	-.001**	-.004*	-0.005*	-.011*
selfemploy_lag	.005	-.007	-.001	.003	-.002	.002	.001	.0004	-.004
unemployed_lag	.029*	.018*	.025*	.023*	.012*	.006*	.022*	.0174*	.035*
retired_lag	.046*	-.004	.015*	.012*	0.012**	.007*	.041*	0.03*	.053*
housework_lag	-.004	.005	.017*	.043*	.002	.006*	.032*	.022*	.029*
inactive_lag	.043*	.014*	.043*	.065*	.076*	.032*	.154*	.086*	.096*