The causal effect of income on health: An original microsimulation approach

Amélie Adeline*
Eric Delattre†

November 25, 2016

Work in progress. Please do not cite or quote without the authors' permission.

Abstract

*ThEMA, University of Cergy-Pontoise. Email: amelie.adeline@u-cergy.fr
†ThEMA, University of Cergy-Pontoise. Email: eric.delattre@u-cergy.fr

SHARE Project acknowledgement - Research Data Center and Data Access

"This paper uses data from SHARE Waves 1, 2, 4 and 5 (DOIs: 10.6103/SHARE.w1.260, 10.6103 - SHARE.w2.260, 10.6103 - SHARE.w4.111, 10.6103 - SHARE.w5.100), see Börsch-Supan et al. [2013] for methodological details.

The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N 211909, SHARE-LEAP: N 227822, SHARE M4: N 261982). Additional funding from the German Ministry of Education and Research, the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11, OGHA 04-064) and from various national funding sources is gratefully acknowledged (see www.share-project.org)."

Contents

Introduction

The relationship between health and individual income is a topic at the center of health economics, with the broad finding that higher socioeconomic status is associated with better health (Preston [1975]). This relationship has been reviewed in many countries and concerning a wide variety of health outcomes (see for instance Wilkinson et al. [1996]; Van Doorslaer et al. [1997]; Kawachi and Kennedy [1999]). While this relationship appears to be well-known, this is not always true concerning its causal interpretation. There are many possible pathways through which earnings fluctuations can impact health. Indeed income, or more generally socioeconomic status is correlated with health in a causal way of the former on the latter. However, we can also think of the reverse association by stating that poor health status may influence income, by reducing the ability to work (Apouey and Clark [2015]). This lack of a clear understanding is an important omission and the direction of the impact of an income shock on health does not seem obvious. Policy makers who aim at improving general health or narrowing health inequalities in a society, need to consider the true direction of causality between income and health. In the absence of randomized controlled experiments, which are not really feasible in this context, the difficulty in disentangling causes and effects is due to endogeneity. Indeed, if the health and the income determine each other simultaneously, then there is an endogeneity issue in their relationship. As a result, since a simultaneous causality in both directions may exist, testing the causal impact of income on health implies to assess the exogeneity of income.

Moreover, identifying the factors that influence the age profile of the self-perceived health is a difficult undertaking, which justifies further investigation. In this way, we intend to do an original microsimulation method (following the methodology of Dormont et al. [2006]) in order to analyze health changes over time. Thanks to this method, we will be able to separately identify changes which are due to changes in morbidity or due to the age of individuals on one hand, and to others changes due to individual characteristics on the other. Concerning morbidity, we will consider a vector of chronic illnesses and disability indicators. This microsimulation method apprehends to identify the impact of income changes while controlling for the age, the morbidity and the technological progress on the self-perceived health status.

To correctly identify these factors, one need to control for their possible endogeneity towards the self-perceived health status. Indeed, individuals take into account several elements of their health and transcribe them into this subjective measure of health. Functional limitations for which an individual is treated, diseases, diagnosed health problems as well as interactions with health professionals are factors which influence the self-rated health (Tubeuf et al. [2008]). This measurement even if it is subjective, is a good predictor of an individual's health (Benitez-Silva et al. [2004]). Thus, it interprets factors which are not always observe by health professionals since it integrates personal expectation of a level of health.

This paper adds a contribution to these subjects by estimating changes in the self-perceived health status following an income shock. As a result, it contributes to the literature concerning causality issues since we control for some aspects that could influence the causal link using an original microsimulation approach. Moreover, we use an instrumental variable approach as well as exogenous shocks in order to get rid of the endogeneity issues related to income.

In section 1 we present the theoretical framework of the causal relationship between income and health. Section 2 describes the econometric framework as well as the microsimulation approach. Then, in section 3 we detail our data. Section 4 reports on the results of the empirical analysis. Section 5 concludes the paper.

1 The causal relationship

1.1 Literature review

The relationship between self-perceived health status and individual income is a heavily researched topics in health economics and other social sciences. Moreover, recent studies model the dynamics of this relationship and question the existence of a causal effect of income changes on health (see for instance Ettner [1996] ; Meer et al. [2003] : Frijters et al. [2005] ; Gardner and Oswald [2007] ; Adda et al. [2009] or Apouey and Clark [2015]). The direction of the causality is considered as an important issue and this is debated among economists since the lack of a clear and true understanding is a shortcoming for policy makers who aim at narrowing health inequalities and thus improving health. The difficulty in disentangling causes and effects is due to endogeneity since in this context, researchers are not able to run randomized controlled experiments.

In this paper, on one hand we want to investigate the direction of the causality by tackling the question of what happens to a person's health when they experience a shock to their income. On the other hand, in the causal relationship between health and income, there are likely to have some effects which need to be controlled.
On the graph 1, we notice that the health status is a decreasing function of age. When people get older, they consider themselves less healthy. Changes in the health status are partly due to the age. As a result, researchers should control for this factor when they want to establish a causal link between income and health. We also see that the dotted lines are almost identical for the four time-periods nevertheless all four lines are not placed on top of each other. This can be due to changes in behaviors on one hand and changes in age, morbidity or technological progress, on the other. These last two effects are likely to be important explanatory factors in the relationship but they might be endogenous to the self-perceived health status. Indeed, self-rated health assimilates the morbidity which depends on diagnosed health problems, interactions with health professionals as well as diseases and functional limitations for which an individual is treated (Tubeuf et al. [2008]).
Self-perceived health status assesses the general perceived health of an individual. This indicator, despite its subjective nature, is a good predictor of people's health status (Benitez-Silva et al. [2004]; DeSalvo et al. [2005] ; Bond et al. [2006]). In order to collect this information, individuals have to choose between five answer categories. On one hand two are at the lower end of the scale ("fair" and "poor"), and on the other hand, two are at the upper ("very good" and "excellent"). The intermediate outcome can be categorized as a neutral term (good, ie. nor very good, nor poor). Thus, it is an important predictor of an individual's health since it comes from different elements that an individual knows about his health. This

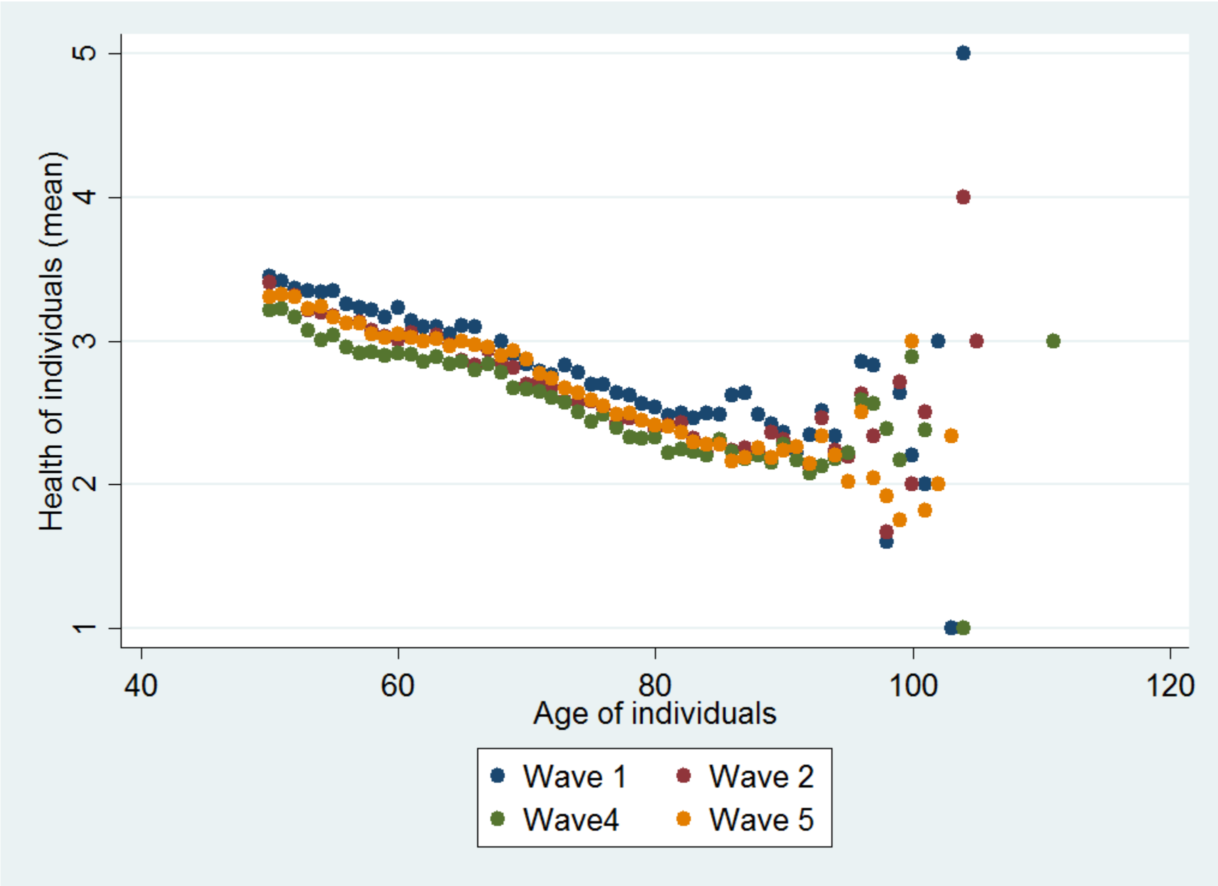

Figure 1: Health of individuals - SHARE survey

subjective measure also integrates factors which are not always considered by health professionals. Finally, it includes individuals' beliefs and attitudes towards a health commodity.

Traditional measures of morbidity provide information about levels of health. The morbidity corresponds to the incidence of diseases. It seems that morbidity is a good predictor of self-assessment of health status and this is why we will control for this effect in the relationship. Indeed, when assessing the health gap between individuals, it is important to measure health status in terms of non-fatal health outcomes. We model the morbidity thanks to an indicator characterized by chronic illnesses and disability indicators.

Then, the last impact we should be careful about is the technological progress. Indeed, the self-perceived health status is expected to increase across-the-board in the future thanks to technological and societal trends which allow an improvement in medicine. Technology has nestled into daily life at a rapid pace. In the "sixth edition of the Public Health Status and Foresight Report" about a "Healthier Netherlands" (Hoeymans et al. [2014]), the authors explain the impact of technological progress. It is said that technological applications are seen in prevention, treatment and care.

Their benefits range from improved diagnostic skills to regenerative medicine to facilitate the independent living. For example, researches enable more targeted prescription of medicines and, sensor technology enables instruments that monitor health status and home automation devices. Furthermore, technological trend can be modelled with a proxy about the longevity. Data on death and its causes are vital measures of a population's health. Examining trends and patterns in mortality can help to explain changes and differences in health status, and evaluate health strategies. Besides, technological trend can also be modelled using a variable which is homogeneous across individuals for a given year. The pattern of these causes reflect changes in behaviors as well as the effects of medical and technological advances.

Indeed, in order to establish a "pure and true" causal relationship between health and income, the goal of this paper is to control for these effects.

1.2 Endogeneity issues

To formulate a causal relationship between self-perceived health status and income, one has to be careful about the endogeneity issues which can arise. Indeed, as explained earlier, we can think of diverse direction of the causality between health and income. In order to assess the real impact of income on health, we focus on the concept of Granger causality. The definition of causality by Granger [1969] distinguishes lag causality from instantaneous causality. As a result, we are going to investigate the causal impact of past income on current health status. This approach also includes a phenomenon of persistence for the health status in the relationship. Moreover, one should also be careful towards the endogeneity of the morbidity indicators. Thus, we intend to estimate the following health equation:

$$\forall i = 1...N \quad \& \quad \forall t = 1...T_i$$

$$h_{it} = \alpha_0 + \lambda h_{i,t-1} + \delta inc_{i,t-1} + X_{it}\beta + Z_{it}\Omega + c_t + \epsilon_{it} \quad (1)$$

where T_i corresponds to the number of observations for an individual i; h_{it} is the self-perceived health status of individual i at date t; inc_{it} is the income of individual i at date t; c_t represents the technological trend and ϵ_{it} is an error term. We consider two sets of observed variables X_{it} and Z_{it} for which we have to control for their exogeneity in this relationship. Thus, they can be either exogenous or endogenous depending on their respective impacts on health. These variables represent the morbidity indicators, the age, the gender, the marital status as well as the schooling. Hereafter, we detail our approach concerning the endogeneity of the different variables.

The first thing to notice is the auto-regressive form of the equation 1 which comes from the fact that we want to highlight the Granger causality of income on health. This auto-regressive form implies a biased estimation if we have:

$$E(h_{i,t-1}.\epsilon_{it}) \neq 0 \quad \forall t \quad \text{if} \quad \epsilon_{it} \neq WN$$

where WN stands for the white noise of an auto-regressive process. However, here we assume that ϵ_{it} is a white noise such that:

$$E(h_{i,t-1}.\epsilon_{it}) = 0 \quad \forall t$$

There is no endogeneity issues if the following assumption holds true, and thus our estimation will not be biased.

Concerning the income variable, the Granger causality involves a delayed causality of income on health in a manner that the income creates disparities all the time. In other words, the lagged income always has an impact on current health such that a permanent variation in individuals income will have a permanent impact on health. Moreover, the income implies the health and might also implies other unobservable variables (such as the life-style of the food expenditures) which in turn might influence the health status. The income variable might be endogenous into the relationship, implying the following :

$$E(inc_{it'}.\epsilon_{it}) \neq 0 \quad \forall t' \leq t$$

As a result, in order to solve this endogeneity issue, we are going to implement an instrumental variables approach as well as exogenous income shocks to the estimation. In the case of endogeneity, standard estimation procedures which assume that income is exogenous will produce biased estimates of the parameters. The instrumental variables approach allows the researcher to implement a consistent estimation method. To use this method, we need a variable, x_{it}^k which is not in equation 1 and which satisfies two conditions:

1. x_{it}^k must be uncorrelated with ϵ_{it} in order to be exogenous: $Cov(\epsilon_{it}, x_{it}^k) = 0$;
2. x_{it}^k is correlated with the income variable: $Cov(x_{it}^k, inc_{i,t-1}) \neq 0$.

When x_{it}^k satisfies the two above conditions, then it is said to be an instrumental variable candidate (or an instrument) for the endogenous variable. By using this method, we intend to correctly identify the causal relationship between the self-perceived health status and the income.
In the literature, some papers use instrumental variables methods or exogenous income shocks to set up a causal link between health and income. Ettner [1996] examines the effect of income on different health proxies, such as self-assessed health,

daily activity limitations, proxies for alcohol abuse and others. Ettner uses cross-sectional data from a number of US surveys collected in the 1980s. Depending on the health outcome, she uses ordered probit, dichotomous probit or two-parts models. She addresses the problem of reverse causality via instrumental variables method, using parental education, work experience, spousal characteristics and unemployment rate as instruments. However, this is a strong assumption because employment seems to have a direct effect on health (unemployed individuals have lower psychological health independent of their lower income). It can be proved that her instruments are not really exogenous (see Meer et al. [2003]). In each case, Ettner finds that income still has a significant impact on health.

Meer et al. [2003] also use instrumental variables method in order to deal with the endogeneity issue on American data. As income variable they use change in wealth, which is instrumented by the amount of inheritances and gifts received over the last five years (amount larger than US $10,000). In the instrumental variable estimation, wealth does not have a significant effect on health. Moreover, the validity of inheritance information as an instrument is also open to debate as noticed by the authors.

Lindahl [2005] uses Swedish longitudinal data to account for the relationship. In this paper, lottery prizes are used to provide exogenous variations in income. However, the identification of lottery winners is not ideal since it is not possible to establish when the individual wins in his life. He runs the estimation on different aspects of health. Lindahl finds that lottery winnings have a positive impact on mental health and it implies a lower body mass index. However, lottery winnings have no effects on some physical health problems. Thus, Lindahl's results are varied.

Frijters et al. [2005] analyzes the association between self-assessed health and income using German data. Their instrument method is to use an exogenous change in income due to the fall of the Berlin wall. In other words, they investigate whether there was a causal effect of income changes on the health satisfaction of East and West Germans in the years following reunification. Results prove a positive impact of income on health.

Gardner and Oswald [2007]'s study explores the causality issue using medium-sized lottery wins (£1000+) as their instrument with British data. They use medium-sized lottery wins because individuals who get no win are almost indistinguishable in their responses from individuals with a small win. They find that mental health is positively affected by income in this case.

Work by Adda et al. [2009] has modelled income and health as a stochastic process which evolves over the life cycle by creating a synthetic cohort dataset, based on successive years of micro data from several English cross-sectional surveys. They exploit the fact that at the cohort level, over the eighties and nighties, there has been sizeable changes in income mainly due to changes in macroeconomic environ-

ment. Their results imply that income variations have little effects on health, but do affect health behaviors and mortality.
Moreover, some researchers also investigate the impact of household income on children's health by using instrumental variables approach to get rid of the endogeneity of income (Case et al. [2002]; Kuehnle [2014]). They all instrument the income using labor characteristics of workers in the households on US and UK datasets. The results are diverse depending on the database.
Last, Apouey and Clark [2015] determine the exogenous impact of income on different health outcomes using lottery winnings to make causal statements (English data). They find that positive income shocks do not have a significant effect on general health, but do have an effect on mental health.

Then, in the estimation of equation 1 we should control for the exogeneity status of the two sets of variables (X_{it} and Z_{it}). We decide to separate the morbidity indicators from the other variables in order to see the specific impact of the morbidity on health. Thus, we consider the exogeneity of what we are calling the covariates hereafter (the age, the gender, the marital status and the schooling) such that:

$$E(X'_{it}.\epsilon_{it}) = 0 \quad \forall t$$

The only role of the above covariates is to control for effects across individuals other than the income, the morbidity, the age or the technical progress which influence the self-perceived health. We thus will not apply a causal interpretation to their coefficients.

However, we need to focus on the possible endogeneity of the morbidity indicators which can biased our estimation. All the morbidity indicators are self-reported and therefore, essentially subjective. Indeed, there might be an instantaneous correlation such that:

$$E(Z'_{it}.\epsilon_{it}) \neq 0 \quad \forall t$$

Indeed, we can think of the following example: if an individual is hit by a car and he breaks his leg following this accident. This may limit him in usual activities. Thus, he will answer positively to the fact of being limited due to health problems (part of the morbidity indicators), impacting his self-perceived health status as well. However, once he will recover, he might feel better and don't be limited any more. Moreover, Cabrero-García and Juliá-Sanchis [2014] explain that "the greater the reported morbidities, the more limited is the subject's activity and the poorer his health". We intend to get rid of this bias with our microsimulation approach in which we control for each factors which can influence the health status. Indeed, a continuous health status comes out from this method which is derived from the lagged value of each factors. Thus, when we consider the lagged values of the morbidity indicators, we no longer have this instantaneous correlation which makes our estimation biased due to the endogeneity of morbidity indicators since the error term will not be the same after the microsimulation approach.

Finally, in this paper we choose to only focus on the instrumentation of the income and not on the one of the morbidity. Indeed, we think that morbidity only have an instantaneous endogeneity effect on self-perceived health, implying that with lagged values of morbidity, we no longer have this endogeneity issue. Moreover, finding good instruments is a complicated task, thus since we focus on the health and income relationship, we concentrate ourselves on the research of appropriate instruments for income only.

2 Econometric framework

The objective of this paper is to formulate a causal relationship between self-perceived health status and income. As a result, two sub-objectives have to be highlighted. The first one is the establishment of a causal link by controlling for the age and other factors that can influence the health, using the concept of Granger causality. The second objective is to assess the endogeneity of income with instrumental variables techniques and by applying exogenous income shocks to the relationship.

2.1 Microsimulation approach

Microsimulation models are useful to establish the effectiveness of health policies in order to understand their value in improving health and to reduce inequalities in health. Accounting for issues such as population heterogeneity and the capacity to capture the long run effects of an intervention are challenging for the identification of the effects of a policy that hindered traditional methods for policy evaluation. An advantage of microsimulation models is that they correspond to an ex-ante evaluation which has the ability to predict the potential impact of a specific policy under different scenarios. There exists different types of microsimulation model (see for example, Zucchelli et al. [2012] for a review of these models). In this paper, we attend to settle our own microsimulation method following the method of Dormont et al. [2006].

The first step of our microsimulation will be to estimate the health equation 1. We will identify the impact of each factors by using the estimated coefficients obtained from the regression. Indeed, in order to do this simulation we need to estimate the model explaining the self-perceived health status. We use an ordered probit model since our variable of interest was initially qualitative, then transformed into a categorical variable by us. When the self-perceived health status outcome is denoted as h_{it}, the model can be stated as

$$h_{it} = j \quad \text{iff} \quad \mu_{j-1} < h^*_{it} \leq \mu_j, \quad \text{for} \quad j = 1, 2, 3, 4, 5$$

The latent variable specification of the model that we estimate can be written as:

$$h^*_{it} = x_{it}\beta + \epsilon_{it}$$

where h^*_{it} is a latent variable which underlies the self-reported health status[1]; x_{it} corresponds to the variables of equation 1; and ϵ_{it} is an error term, which can be decomposed into two terms $\eta_i + \zeta_{it}$ and assumed to be normally distributed. In this research, the latent outcome h^*_{it} is not observed. Instead, we observe an indicator of the category in which the latent indicator falls. As a result the observed variable is equal to 1, 2, 3, 4 or 5 for "fair", "poor", "good", "very good" or "excellent" with these probabilities:

$$P(y = j|x) = F(\mu_j - x_i\beta) - F(\mu_{j-1} - x_i\beta)$$

where the interval decision rule is:

1. $h_{it} = 1$ if $h^*_{it} \leq \mu_1$;
2. $h_{it} = 2$ if $\mu_1 < h^*_{it} \leq \mu_2$;
3. $h_{it} = 3$ if $\mu_2 < h^*_{it} \leq \mu_3$;
4. $h_{it} = 4$ if $\mu_3 < h^*_{it} \leq \mu_4$;
5. $h_{it} = 5$ if $h^*_{it} > \mu_4$.

In this model, the threshold values $(\mu_1, \mu_2, \mu_3, \mu_4)$ are unknown. We do not know the value of the index necessary to go from very good to excellent. In theory, the threshold values are different for everyone.

Running the model with composite error will lead us to estimate the coefficients of equation 1. Moreover, we will control for the age, the technological progress and the morbidity to identify their specific contributions to the health of individuals.

As explained above, the starting point of our microsimulation approach consists in estimating equation 1 with an ordered probit model in order to get the contributions of each factors to the health of individuals. Once we get this estimation, we can focus on the changes in health while controlling for some factors. The different steps

[1] Once h^*_{it} crosses a certain value you report fair, then poor, then good, then very good, then excellent health.

of the simulation are the following:

1. Once we get the estimated coefficients, the next step of the microsimulation allows to control for the age of individuals.

$$\tilde{h}^1_{it} = \lambda h_{i,t-1} + X_{it}\hat{\beta} + \hat{\delta} inc_{i,t-1} + Z_{it}\hat{\omega} + \hat{c}_t \quad (2)$$

This estimate informs us on the impact of all factors on the self-perceived health status of individuals in t, except the age. Indeed, we estimate the health in t with the age of individuals in $t-1$, thus we can see what are the different impacts on the health while keeping the age of individuals constant. As a result this gives us the effects that are not due to the age.

2. Then, the estimate highlighting the changes in health which are due neither to the age nor to the morbidity, can be written as:

$$\tilde{h}^2_{it} = \lambda h_{i,t-1} + X_{it}\hat{\beta} + \hat{\delta} inc_{i,t-1} + Z_{i,t-1}\hat{\omega} + \hat{c}_t \quad (3)$$

where Z_{it} is replaced by $Z_{i,t-1}$ in equation 2.
In this equation, we now control for two effects which can influence the relationship: the age of individuals and the morbidity which corresponds to the prevalence of diseases. The difference between the estimated coefficients of equations 2 and 3 allows us to see the impacts of the factors on health, given the age and the morbidity.

3. The penultimate impact we want to control for is the impact of the technological trend on health. This can be seen through the following equation:

$$\tilde{h}^3_{it} = \lambda h_{i,t-1} + X_{it}\hat{\beta} + \hat{\delta} inc_{i,t-1} + Z_{i,t-1}\hat{\omega} + \hat{c}_{t-1} \quad (4)$$

where c_t is replaced by c_{t-1} in equation 3.
This equation gives the changes in the self-perceived health status while controlling for the age, the morbidity as well as the technological trend. Technological trend is a factor which is homogeneous for all individuals for a given year.

4. Finally, in order to obtain the final equation of the microsimulation we have to control for the individual effects of socioeconomic characteristics, in other words individual behaviors (Dormont et al. [2006]):

$$\tilde{h}^4_{it} = \lambda h_{i,t-1} + X_{i,t-1}\hat{\beta} + \hat{\delta} inc_{i,t-1} + \hat{\omega} Z_{i,t-1} + \hat{c}_{t-1} \quad (5)$$

where X_{it} is replaced by $X_{i,t-1}$ in equation 4.
In this equation, we add a control for the information on individuals demographic characteristics. Individual characteristics correspond to the gender, the schooling and the marital status. First, concerning the gender, this is fixed across the waves. Then, for the other variables, each component provides different resources and dis-

plays different relationships to various health outcomes. As a result, concerning the schooling, a higher level will allow an individual to have better access to health systems and therefore its subjective health should be better. Education shapes future occupational opportunities and earning potential. Thus, it also provides knowledge that allow better-educated persons to gain more ready access to information which promote health. Moreover, Grossman [1972] considers that variables such as the age and the education will influence the optimum level of health. As a result, if one decides to control for the age then we should also control for the education. We consider these individual socioeconomic characteristics as exogenous since their roles in the estimates are to be variables of control.

Thus, taking into account of all these factors separately helps us to establish a real causal link between health and income. Indeed, the goal of this paper is to establish a causal relationship between self-perceived health status and individual income without the effect of the increase in the age of individuals as well as changes associated to the others variables. After this last step we end up with a continuous health variable which can be used to re-estimate the relationship using an ordinary least square method:

$$\tilde{h}^4_{it} = \kappa_0 + \lambda h_{i,t-1} + X_{it}\beta + \delta inc_{i,t-1} + \omega Z_{it} + c_t + \nu_{it} \qquad (6)$$

where κ_0 is a constant term and ν_{it} is the error term.
The microsimulation approach allows us to get rid of the endogeneity issue associated to the morbidity indicators since $\tilde{h}^4_{it}$ does not include unobservable components which might explain the morbidity in t and the health status in t. This health status is derived from the lagged explanatory variables, including the lagged morbidity indicators, such that we have the following:

$$E(Z_{it}.\nu_{it}) = 0$$

This enables us to go further in the establishment of the causality of income on health. Moreover, as our principle objective is to assess the causal impact of income on health, we focus our study on the instrumentation of income and not on the one of morbidity. Besides, finding instruments for the morbidity is a complicated task and researchers should be conscientious in the choice of instruments. As explained earlier, we hope and think to get rid of the endogeneity issue associated to the morbidity indicators since we introduced their lagged values in our microsimulation approach.

2.2 Instrumentation of the income

In the health economics literature about causality, the difficulty to distinguish the causes and the effects is due to endogeneity issues. Wooldridge [2002] brings two issues to the forefront which need to be taken into account to solve the endogeneity problem:

1. The issue of reverse-causality is a concern when one studies the income-related health relationship: a positive income shock can lead to an improvement in the health status through, for example, better access to medical services. However, we can also think of the reverse relationship explaining that people in good health are likely to be more economically productive and thus have higher incomes.
2. Some individual characteristics which are not identified by the researcher may determine both income and self-assessed health status. A biased estimation between income and health can result from a failure of controlling for these effects.

From an early stage in the debate, it was argued that higher income causes better health (Preston [1975]). Smith [1999] explaines that this positive relationship leads to a number of interpretations: causality may goes from income to health (high economic resources lead to better health status for many reasons such as: more resources devoting to health or better knowledge about what improves health), from health to income (poor health may restrict a family's capacity to earn income or to accumulate assets by limiting work or by raising medical expenses), or both may be determined by other common factors. There have been several attempts to deal with the problem of endogeneity. The idea is to find a variable x_{it}^k which is correlated to the endogenous variable $inc_{i,t-1}$ but which is not correlated with the error term ν_{it}. Indeed, when one of the coefficients (e.g. δ), which defines the relationship between a variable (e.g. the income $inc_{i,t-1}$) and the dependant variable (in our case the health h_{it}) cannot be interpreted in a causal way, this might be because of endogeneity issues of the income variable. As explained earlier, this issue implies:

$$E(\nu_{it}|inc_{i,t-1}) \neq 0$$

This can biased the estimation and the interpretation. Indeed, if health and income are simultaneously determined, the endogeneity problem implies that standard estimation procedures which assume that income is exogenous will produce biased estimates of the parameters. As a result, one has to find some solutions like instrumental variables method or the use of exogenous shocks on the endogenous variable, to solve this issue. In section 1.2 we review the recent papers which attempt to establish the causality between income and health. These papers try to produce consistent estimates of the effect of income on health using instrumental variables techniques or exogenous income shocks. In the case where the relationship is not exogenous then the estimation will not be convergent (Dormont [2007]). Indeed, we need to estimate a model representing a causal relationship (Goldberger [1972]), contrary to a model which only highlights a relationship that simply captures statistical associations. The instrumental variables approach allows us to eliminate all biases due to the correlation of income with the error term in the health equation such that we have the following estimate:

$$\tilde{h}^4_{it} = \kappa_0 + \lambda h_{i,t-1} + X_{it}\beta + \delta inc_{i,t-1} + \Theta x^k_{it} + Z_{it}\omega + c_t + \nu_{it} \tag{7}$$

which consists in the equation 6 with x^k_{it} the instrument we introduce.
This equation corresponds to the first way of assessing the endogeneity issue since it employs an instrumental variables technique (Θx^k_{it}). Instrumenting eliminates measurement errors as well as endogeneity bias. The data availability allows us to choose between different instruments. As a result, we explain our choice hereafter. At a macroeconomic point of view, we can use the unemployment rate of each country and each wave since this is correlated to the amount earned each month. Whether the individual has an income also depends on whether he is working. However, this will be a valid instrument only if the changes in health are due solely to differences in income. At the microeconomic level, we can also use the location of the main residence of individuals as a valid instrument for the income. Indeed the income is correlated with where individuals live. We can assume that the location of the main residence is correlated to the employment areas, and thus to the income of individuals. In order to see if these variables are good instruments, we perform the first stage estimation (see table 12 in the annexe section). The results of this first stage shows that all these variables are statistically significant, or in other words they are good predictors of the income.

Besides, to be sure of assessing correctly the endogeneity issue we also apply an exogenous income shock to the equation 7. Thanks to the data availability, we can follow the intuition of Meer et al. [2003] using an information about the amount of financial or material gift received (worth 250€ or more) and the amount of gift or inheritance (worth 5 000€ or more). These two variables correspond to unexpected gifts or inheritances which are assumed not to be endogenous. These information will be included as dummy variables so that the final equation to estimate is:

$$\begin{aligned}\tilde{h}^4_{it} = \kappa_0 + \lambda h_{i,t-1} + X_{it}\beta + \delta inc_{i,t-1} + \Theta x^k_{it} + \Gamma_1 \mathbb{1}_{GIFT_1} \\ + \Gamma_2 \mathbb{1}_{GIFT_2} + Z_{it}\omega + c_t + \nu_{it}\end{aligned} \tag{8}$$

which is an extension of equation 7 where Γ_1 and Γ_2 corresponds to the impact of the dummy variables associated to the exogenous income shocks.

3 Data

The complex interaction of changing humans in changing environments is not thought to be captured adequately by simple relationships among variables at a point in time and this is why we want to explore the panel dimension of a database.

3.1 SHARE Survey

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of more than 123 000 individuals aged 50 or over from many European countries and Israel. Since 2004, SHARE asks questions throughout Europe to a sample of households with at least one member who is 50 and older. These households are re-interviewed every two years in the panel. The first wave (2004-2005, 27 014 individuals) and the second one (2006-2007, 34 393 individuals) were used to collect data on health status, medical consumption, socio-economic status and the living conditions. The 2008-2009 survey (Wave 3) "SHARELIFE" was extended to life stories by collecting information on the history of the respondents. Since it does not contain the required information for our research, this wave is not taken into account in the pooled database used. The number of participants increased from 12 countries in wave 1 (Börsch-Supan [2016a]), to 15 (+ Ireland, Israel, Poland and Czech Republic) in wave 2 (Börsch-Supan [2016b]), and the third wave contains information about 13 countries. The wave 4 (2010-2011), is a return to the initial questionnaire waves 1 and 2 (Börsch-Supan [2016c]). It collects data from 56 675 individuals aged 50 and over in 16 European countries. Finally, the fieldwork of the fifth wave (Börsch-Supan [2016d]) was completed in November 2013. The following countries are included in the scientific release of 2015: Austria, Belgium, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, France, Israel, Italy, Luxembourg, Netherlands, Sweden, and Slovenia. This wave contains the responses of 63 626 individuals. As a result, the pooled database contains information on 180 606 individuals which are present about 1.7 years in the panel. However, researchers should be also aware of the potential disadvantage of this database. Indeed, Börsch-Supan et al. [2013] explain that in some waves there are a relative low response rates and moderate levels of attrition (even though the overall response rate is high compared to other European and US surveys with an average retention rate over the year of 81 %) which are due to the economic crisis faced by some countries implying a decrease in the participation rates.

We choose to focus on this survey since it has all the information needed to carry out this research. Indeed, the dependant variable in our study is the self-perceived health status where individuals are asked to classify their health from "poor" to "excellent" (see figure 2).

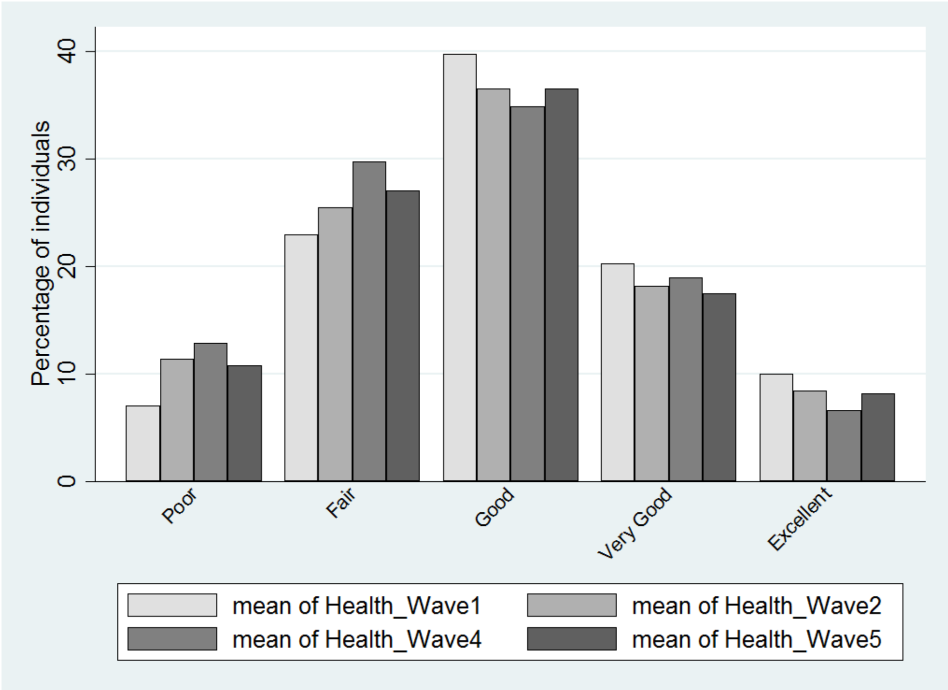

Figure 2: Distribution of the self-perceived health status - SHARE

3.2 The income variable

As explained earlier in this paper, we have to control for the income. In this database, the income corresponds to the sum of individual imputed income for all household components. We intend to apply two methods to make sure of the robustness of the causal link. First, we chose to apply an instrumental variables method to get rid of the endogeneity issue. As mentioned earlier, a good instrument has to be correlated with the income but not with the health of individuals. Due to data availability, we decide to introduce into the econometric analysis, two variables which have been previously estimated and chosen from an auxiliary equation in a first stage (see tables 8 to 12 for detailed statistics). The first variable is a microeconomic instrument corresponding to the location of the main residence. This is a categorical variable in which individuals say whether they live in a big city, the suburbs or outskirts of a big city, a large town, a small town or a rural area. We thus create dummy variables and take as a reference category to the interpretation "living in a rural area". The second variable is the unemployment rate[2] and can be

[2]Source: OECD website.

considered as a macroeconomic instrument since it is computed for each country at each year of the survey (Meer et al. [2003]). Second, we would like to know what are the changes in the health status following positive income shocks. As a result, we have to find a variable implying a positive shock of income for individuals. The fact of using data on inheritance creates a setting as close as possible to the idealized laboratory experiments. This SHARE panel database allows researchers to choose two variables. Indeed, one information given in the survey is on whether an individual has received an amount of 250€ or more (material gift or not) or if the individual has ever received a gift or inheritance (worth 5 000€ or more). Moreover, we can ask whether inheritance/financial gift is a good instrumental variable. First of all, we are mindful of the possible concerns with our instrumentation strategy. Indeed, we can argue that inheritance does not satisfy all the exclusion restrictions in a way that this is not a very strong income shock (a family member dying might signal something about the individual's health or unobserved variable might drive both heath and inheritance with the idea of "privileged backgrounds"). However, the significant power of these variables have been tested and validated from a first stage regression (12 in the appendix part).

3.3 Measurement of the morbidity

It is important to measure health status in terms of "non-fatal" health outcomes since they are important concerning the burden of a disease. Morbidity indicators can be broadly defined by the prevalence or incidence of diseases but also by the degree of disability and the risky behaviors of individuals which can cause diseases. Morbidity is strongly correlated with the self-perceived health status (Manor et al. [2001]; Latham and P. [2012]; Chan et al. [2015]). As a result, it has to be taken into account when one studies the self-perceived health status. With this measure we would like to know if an individual, given his/her age, is more often affected by a disease in $t+1$ than in t.

Dormont et al. [2006] use a microeconomic dataset which is a survey (Santé Protection Sociale) conducted by IRDES (Institute for Research and Information in Health Economics, Paris) in order to construct morbidity indicators. We will base our construction of indicators on their method since they produce these indicators with the help of general practitioners to make sure of the validity of the latter. As regards morbidity, we consider the last two indicators of the Mini European Health Module (MEHM). The MEHM is included in several European survey programs including the EU-SILC, SHARE, EHIS and Eurobarometer. It consists in indicators representing three concepts of health. The first one concerns the self-perceived health status which assesses general perceived health rather than the present state of health. This indicator first recommended by the World Health Organization in

1988, seeks to incorporate different dimensions of health (i.e. physical, social, and emotional as well as functional signs and symptoms). Despite its subjective nature, indicators of perceived general health have been found to be a good predictor of people's future health care use and mortality (DeSalvo et al. [2006]; Cox et al. [2009]). The second indicator is the morbidity and it assesses the incidence or prevalence of a disease or of all diseases. This indicator gives information about people having long-standing illness or health problems. The last one is about activity limitation and disability which assess self-perceived long-standing limitations in usual activities due to health problems.
Thus we will use a vector of chronic illnesses and disability indicators for the morbidity. Indeed, a variety of lifestyle factors and health-related behaviors such as alcohol consumption, physical activity and dietary habits can affect a person's health. An unhealthy lifestyle often results in a higher risk of chronic diseases.

Our database (the SHARE survey) presents the advantage of providing information about many morbidity indicators which can be distinguished into three main parts[3]. The first one concerns the degree of invalidity of individuals. In this category, we have four indicators. Activities of Daily Living (ADLs) consist in "basic activities that are necessary to independent living (e.g. walking, bathing, dressing, toileting, brushing teeth and eating)", according to the World Health Organization (WHO). This concept determines an individual's ability to perform the activity with or without assistance. Instrumental Activities of Daily Living (IADLs), according to the World Health Organization, are "activities with aspects of cognitive and social functioning, including shopping, cooking, doing housework, managing money and medication, and using the telephone or the computer". These tasks support an independent lifestyle. The Global Activity Limitation Indicator (GALI) is part of the MEHM. It belongs to the family of disability indicators, targeting situations in which health disorders and conditions have impacted people's usual activities (number of limitations with mobility, arm function and fine motor). It is a single-item survey instrument where individuals are asked: " For at least the last 6 months, have you been limited because of a health problem in activities people usually do?" and they have to answer using: "1) Yes, strongly limited; 2) Yes, limited; 3) No, not limited". The second indicator is about chronic diseases. Chronic diseases give the number of chronic diseases of an individual (heart problem, high blood pressure/high blood cholesterol, a stroke or cerebral vascular disease, diabetes, cancer...). Finally, the third category of morbidity's indicators concerns the risky behaviors of individuals. As a result, the drinking variable informs us on the drinking habits. The World Health Organization recommendations for a reasonable consumption is a maximum of two glasses of alcohol per day. However, the WHO

[3]See the appendix part in order to have detailed statistics on the indicators.

also states to abstain from alcohol at least one day in the week, and not to consume more than 4 drinks on one-time opportunity.

3.4 Measurement of the technical progress

The technical progress represents the unobserved heterogeneity. On one hand it can be modelled using, for instance, a proxy for "the lengthening of the lifetime". However, this information is not easily obtainable in the dataset. As a result, a similar information is the life expectancy. The OECD website gives us information about the life expectancy at 65 years old for Europeans countries. We distinguish the women's life expectancy from the men's one in each country in order to have the more accurate information. On the other hand, the technological progress can also be viewed as a variable which is homogeneous for all individuals for a given year. As a result we can also add time's dummy variables in the specification. Since, the life expectancy is not completely collinear to the time's dummy variables, both variables are added into the specification in order to capture the real trend implied by the technical progress.

4 Results

In order to illustrate the usefulness of the microsimulation approach we first present the results of equation 1 where we apply an ordered probit model. Since we want to highlight the Granger causality of income on health, we include lagged variables for the income and the health (phenomenon of persistence). As a result, we lose observations due to these delayed variables as all individuals are not always present during the four waves of the panel. This is why this analysis gives us access to 77 479 observations corresponding to 52 569 individuals. Indeed, at the beginning we have 181 620 observations including 50 972 individuals which are present only once in the panel, 71 674 present twice, 26 154 present during three waves and 32 820 individuals present during the four waves. By adding the lagged values in the microsimulation method we lose 103 732 observations (see table 1). Once we have this information in mind, we can start the econometric analysis. The key variable represents the self-perceived health status of individuals which is recorded in five modalities. The lagged health (which is implied by the Granger causality) shows that there is a strong phenomenon of persistence in the health status. Results in table 2 display that past income have a positive impact on the feeling of individuals about their health today. This result is significant and has the intuitive sign according to the literature where it is said that a higher income improves the health status. Thus, income has a permanent effect on health. The coefficients associated to the age allows us to identify a convex effect of age on the subjective health. This implies that

Table 1: The loss of observations due to lagged variables

Presence of ind. in the panel	Number of obs.	Obs. lost
1 wave	50 972	50 972
2 waves	71 674	÷ 2 = 35 837
3 waves	26 154	÷ 3 = 8 718
4 waves	32 820	÷ 4 = 8 205
Total obs.	181 620	103 732
Nb. obs. after the lagged var.	181 620 - 103 732 = 77 888	
Nb. obs. with the missing values	77 479	

the health is a decreasing function of age and when individuals reached a given age their health cannot be worst. Concerning the morbidity indicators which represent the prevalence or incidence of a disease, the results imply that being affected by a disease or by limitations decrease the self-rated health status. One exception is for the drinking variable. In fact, a received idea among individuals is that drinking two glasses of wine per day decreases the cardiovascular risks. This explains the positive coefficient associated to this variable, meaning that drinking improves their subjective health. However, according to the World Health Organization, this does not decrease risks of cardiovascular diseases and quite the reverse since it is dangerous for the health. Finally for the technical progress, we both include the life expectancy and the cohorts fixed effects (wave 1 is not included since the analysis has been performed using lagged variables). In general, when the life expectancy increases, this will help individuals to feel better.

Table 2: First step of the microsimulation model (equation 1)

Variables	**Coefficients**
Dependant variable: Health_t	
Granger Causality	
Health_{t-1}	0.547 ***
	(0.005)
Log of income_{t-1}	0.086 ***
	(0.003)
Morbidity Indicators	
ADL	-0.055 ***
	(0.008)
IADL	-0.055 ***
	(0.006)
GALI	-0.146 ***
	(0.007)
Mobility indicator	-0.162 ***

Table 2: Results of the first step of the microsimulation model (continued)

Variables	**Coefficients**
	(0.003)
Chronic diseases	-0.195 ***
	(0.003)
Drinking	0.063 ***
	(0.011)
Technical progress	
Wave 2	0.300 ***
	(0.019)
Wave 4	-0.009
	(0.009)
Wave 5	*Reference*
Life Expectancy	0.025 ***
	(0.003)
Co-variables	
Age	-0.032 ***
	(0.006)
Age squared	0.0002 ***
	(0.0001)
Women	0.025 *
	(0.015)
Married	*Reference*
Living with partner	0.015
	(0.029)
Living as a single	0.033 *
	(0.017)
Never married	-0.030
	(0.020)
Divorced	0.018
	(0.016)
Widowed	0.063 ***
	(0.014)
Education	0.006 ***
	(0.001)
Numb. of obs.	77 479
Numb. of groups	52 569

Table 2: Results of the first step of the microsimulation model (continued)

Variables	**Coefficients**
***: 1% significant; **: 5% significant;	
*: 10% significant.	

In order, to have an accurate causal relationship between health and income, one should be careful about the possible endogeneity of the morbidity indicators. Moreover, researchers should also account for some factors which can influence the health-income relationship. As a result, the microsimulation allows to control for all these aspects. This approach allows us to compute a continuous self-perceived health status which is still positively affected by past individual income (see table 3). The signs of the coefficients do not change from the estimation of equation 1 with an ordered probit model to the estimation of equation 6 with an ordinary least square on the pooled database. As a result, we can ensure that our microsimulation approach allows us to compute an accurate continuous health status. Moreover, we get rid of the endogeneity issues of the morbidity indicators on the self-perceived health status by fixing all the morbidity indicators to their lagged values. In this estimation, the morbidity indicators all have a negative and significant impact on the continuous health status, except for the drinking habits which is not statistically significant any more. Concerning the technical progress, the effects are the same as before. This method allows us to correctly identify the effects of the age, the morbidity indicators and the technical progress on the self-perceived health status.

Table 3: Results of the last step of the microsimulation (equation 6)

Variables	**Coefficients**
Dependant variable: Health_t	
Granger causality	
Health_{t-1}	0.718 ***
	(0.002)
Log of income_{t-1}	0.080 ***
	(0.001)
Morbidity Indicators	
ADL	-0.040 ***
	(0.003)
IADL	-0.027 ***
	(0.002)
GALI	-0.048 ***
	(0.003)
Mobility	-0.052 ***

Table 3: Results of the last step of the microsimulation (continued)

Variables	**Coefficients**
	(0.001)
Chronic Diseases	-0.081 ***
	(0.001)
Drinking	0.011 ***
	(0.004)
<u>Technical progress</u>	
Life expectancy	0.025 ***
	(0.001)
Wave 2	0.078 ***
	(0.007)
Wave 4	-0.26 ***
	(0.003)
Wave 5	*Reference*
<u>Covariates</u>	
Age	-0.029 ***
	(0.002)
Age squared	0.0001 ***
	(0.00002)
Gender (=1 if women)	-0.052 ***
	(0.006)
Education	0.006 ***
	(0.0004)
Married, living with spouse	*Reference*
Registered Partnership	-0.019 *
	(0.011)
Married, not living with spouse	0.025 ***
	(0.006)
Never married	-0.015 *
	(0.008)
Divorced	0.001
	(0.006)
Widowed	0.012 **
	(0.005)
Numb. of obs.	77 108
Numb. of groups	52 356

Table 3: Results of the last step of the microsimulation (continued)

Variables	**Coefficients**
***: 1% significant; **: 5% significant;	
*: 10% significant.	

The last step of our study consists in working on the income endogeneity issues. As a result, we decide to both implement an instrumental variables method (two stage least square) as well as adding positive income shocks to the estimation. The estimation of equation 8 is reported in table 4. In this estimation, we instrument the income with the unemployment rate by country and year and the location of the main residence of individuals. The results highlighting the Granger causality are the same as before implying the robustness of our results. Moreover, we also add two income shocks to the estimation. The first one (financial gift of 250€ or more) is not statistically significant such that this does not affect the health of individuals. The second one (financial gift of 5 000€ or more) is statistically significant and implies a negative effect on health. An explanation would be that if an individual, aged 50 and older, inherits this money from a close relation who just died, this will have a negative impact on the health of the individual since he loses someone important (feeling of sadness). Concerning the impact of the morbidity indicators as well as the impacts of the technical progress the results are the same as before.

Table 4: Results with the use of IV and exogenous shocks (equation 8)

Variables	**Coefficients**
Dependant variable: Health_t	
Granger Causality	
Health_{t-1}	0.702 ***
	(0.003)
Log of income_{t-1} *(instrumented)*	0.109 ***
	(0.007)
Exogenous income shocks	
Financial gift (≥250€)	0.003
	(0.007)
Financial gift (≥5000€)	-0.022 ***
	(0.008)
Morbidity Indicators	
ADL	-0.045 ***
	(0.003)
IADL	-0.029 ***
	(0.003)

Table 4: Results with the use of IV and exogenous shocks (continued)

Variables	**Coefficients**
GALI	-0.046 ***
	(0.003)
Mobility	-0.059 ***
	(0.001)
Chronic Diseases	-0.087 ***
	(0.001)
Drinking	0.008 *
	(0.005)
Technical progress	
Life expectancy	0.025 ***
	(0.002)
Wave 2	0.075 ***
	(0.008)
Wave 4	-0.264 ***
	(0.004)
Wave 5	*Reference*
Covariates	
Age	-0.032 ***
	(0.003)
Age squared	0.0002 ***
	(0.00002)
Gender (=1 if women)	-0.045 ***
	(0.007)
Education	0.006 ***
	(0.001)
Married, living with spouse	*Reference*
Registered Partnership	-0.031 **
	(0.014)
Married, not living with spouse	0.037 ***
	(0.007)
Never married	0.005
	(0.009)
Divorced	0.011 ***
	(0.007)
Widowed	0.037 ***
	(0.006)

Table 4: Results with the use of IV and exogenous shocks (continued)

Variables	**Coefficients**
Numb. of obs.	50 333
Numb. of groups	36 238
Instruments: Location of the main residence and unemployment rate.	
***: 1% significant; **: 5% significant; *: 10% significant.	

Conclusion

A heavily researched topics in health economics is the relationship between income and health and more specifically the direction of causality between the two. This paper highlights the question of whether income implies health in a causal way. While it seems well-known that people with higher incomes enjoy better health, it is far more difficult to establish the direction of the causality of this relationship. The definition of causality which is chosen here is the Granger one implying a persistence phenomenon in the relationship. Factors such as the age, the morbidity or the technical progress have to be controlled since they could influence this relationship. We use a rich longitudinal database (SHARE survey) which covers a statistically representative sample of Europeans individuals aged 50 and over and reports detailed information on income and health as well as health behaviors.

The microsimulations performed above enables us to identify the components of the health-income relationship and to control for the endogeneity issues which can arise. Indeed, we assess that this approach allows us to get rid of the endogeneity of the morbidity indicators since we fix these factors to the previous date and look at their impacts on the current health status of individuals. Thus, instantaneous endogeneity of morbidity on health is then no more an issue. Moreover, in order to get rid of the income endogeneity issue which can biased our estimates, we implement an instrumental variables method as well as exogenous income shocks to the estimation.

The results presented here have underlined a central point in the analysis of health economics and income. Researchers need a clear understanding of the direction of the causality in this relationship. Indeed, our method and results ensure the Granger causality of income on health. In other words, we show that income has a permanent effect on subjective health status. We thus get rid of the possible reverse causation in this relationship since our results appear to be robust.

References

J. Adda, H.-M. Gaudecker, and J. Banks. The impact of income shocks on health: evidence from cohort data. *Journal of the European Economic Association*, 2009.

B. Apouey and A. E. Clark. Winning big but feeling no better? the effect of lottery prizes on physical and mental health. *Health economics*, 2015.

H. Benitez-Silva, M. Buchinsky, H. Man Chan, S. Cheidvasser, and J. Rust. How large is the bias in self-reported disability? *Journal of Applied Econometrics*, 19 (6):649–670, 2004.

J. Bond, H. O. Dickinson, F. Matthews, C. Jagger, C. Brayne, and MRC CFAS. Self-rated health status as a predictor of death, functional and cognitive impairment: a longitudinal cohort study. *European Journal of Ageing*, 3(4):193–206, 2006.

A. Börsch-Supan. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 5.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.500. 2016a.

A. Börsch-Supan. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 5.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.500. 2016b.

A. Börsch-Supan. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 5.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.500. 2016c.

A. Börsch-Supan. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 5.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.500. 2016d.

A. Börsch-Supan, M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber. Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology DOI: 10.1093/ije/dyt088*, 2013.

J. Cabrero-García and R. Juliá-Sanchis. The global activity limitation index mainly measured functional disability, whereas self-rated health measured physical morbidity. *Journal of clinical epidemiology*, 67(4):468–476, 2014.

A. Case, D. Lubotsky, and C. Paxson. Economic status and health in childhood: The origins of the gradient. *The American Economic Review*, 2002.

Y. Y. Chan, C. H. Teh, K. K. Lim, K. H. Lim, P. S. Yeo, C. C. Kee, M. A. Omar, and N. A. Ahmad. Lifestyle, chronic diseases and self-rated health among Malaysian adults: results from the 2011 National Health and Morbidity Survey (NHMS). *BMC public health*, 2015.

B. Cox, H. Van Oyen, E. Cambois, C. Jagger, S. Le Roy, J.-M. Robine, and I. Romieu. The reliability of the minimum European health module. *International Journal of Public Health*, 2009.

K. B. DeSalvo, V. S. Fan, M. B. McDonell, and S. D. Fihn. Predicting mortality and healthcare utilization with a single question. *Health services research*, 40(4): 1234–1246, 2005.

K. B. DeSalvo, N. Bloser, K. Reynolds, J. He, and P. Muntner. Mortality prediction with a single general self-rated health question. *Journal of General Internal Medicine*, 2006.

B. Dormont. *Introduction à l'économétrie*. 2nd edition, 2007.

B. Dormont, M. Grignon, and H. Huber. Health expenditure growth: reassessing the threat of ageing. *Health economics*, 2006.

S. L. Ettner. New evidence on the relationship between income and health. *Journal of health economics*, 1996.

P. Frijters, J. P. Haisken-DeNew, and M. A. Shields. The causal effect of income on health: Evidence from German reunification. *Journal of health economics*, 2005.

J. Gardner and A. J. Oswald. Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *Journal of health economics*, 2007.

A. S. Goldberger. Structural equation methods in the social sciences. *Econometrica: Journal of the Econometric Society*, 1972.

C. WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 1969.

M. Grossman. On the concept of health capital and the demand for health. *Journal of Political economy*, 1972.

N. Hoeymans, A.J.M. van Loon, M. van den Berg, M.M. Harbers, H.B.M Hilderink, J.A.M. van Oers, and C.G. Schoemaker. A healthier Netherlands: Key findings from the dutch 2014 public health status and foresight report. *RIVM, Bilthoven*, 2014.

I. Kawachi and B. P. Kennedy. Income inequality and health: pathways and mechanisms. *Health services research*, 1999.

D. Kuehnle. The causal effect of family income on child health in the UK. *Journal of health economics*, 2014.

K. Latham and C. W. P. Self-rated health and morbidity onset among late midlife US adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, page gbs104, 2012.

M. Lindahl. Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human resources*, 2005.

O. Manor, S. Matthews, and C. Power. Self-rated health and limiting longstanding illness: inter-relationships with morbidity in early adulthood. *International Journal of Epidemiology*, 2001.

J. Meer, D. L. Miller, and H. S. Rosen. Exploring the health–wealth nexus. *Journal of health economics*, 2003.

S. H. Preston. The changing relation between mortality and level of economic development. *Population studies*, 1975.

J. P. Smith. Healthy bodies and thick wallets: the dual relation between health and economic status. *The journal of economic perspectives: a journal of the American Economic Association*, 1999.

S. Tubeuf, F. Jusot, M. Devaux, C. Sermet, et al. Social heterogeneity in self-reported health status and measurement of inequalities in health. *Paris: Irdes (12)*, 2008.

E. Van Doorslaer, A. Wagstaff, H. Bleichrodt, S. Calonge, Ulf-G Gerdtham, M. Gerfin, J. Geurts, L. Gross, U. Häkkinen, R. E. Leu, et al. Income-related inequalities in health: some international comparisons. *Journal of health economics*, 1997.

R. G. Wilkinson et al. *Unhealthy societies: the afflictions of inequality.* Routledge, 1996.

J. Wooldridge. Econometric analysis of cross section and panel data. *The MIT Press*, 2002.

E. Zucchelli, A. M. Jones, N. Rice, et al. The evaluation of health policies through dynamic microsimulation methods. *International Journal of Microsimulation*, 2012.

A Descriptive Statistics

A.1 Variables of interest and covariates

Table 5: Descriptive statistics of the variables of interests and some covariates

Variables	Mean	Std. dev.	Min.	Max.	Nb. of obs.
Self perceived health	2.846	1.087	1	5	181 620
Micro-simulated health	-0.195	1.068	-6.625	2.616	77 485
Log of income	9.922	1.304	0.014	16.122	181 620
Exogenous Shock:	%				
Financial gift 250€ or more	7.19				125 483
Financial gift 5000€ or more	16.44				125 170
Instrumental variables:					
Unemployment rate	8.571	4.478	3.2	26.1	181 620
Big city (%)	11.49				
Suburbs of a big city (%)	10.57				
Large Town (%)	13.72				
Small town (%)	18.80				
Rural area (%)	22.48	*Reference*			
Age	66.141	10.086	50	111	181 620
Education:	%				
Without diploma	4.14				
Primary	21.05				
Lower secondary	18.49				
Upper secondary	35.56	*Reference*			
First Stage of tertiary	19.91				
Second stage of tertiary	0.71				
Still in school	0.004				
Missing	0.13				
Marital Status:	%				
Married living with spouse	68.15	*Reference*			
Married living single	9.18				
Registered partnership	2.28				
Never married	3.65				
Divorced	5.79				
Widowed	9.59				

Table 5: Descriptive statistics of the variables of interests and some covariates (continued)

Variables	Mean	Std. dev.	Min.	Max.	Nb. of obs.
Missing	1.35				

A.2 Morbidity indicators

Table 6: Morbidity Indicators

Variables	**Mean**	**Std. dev.**	**Min.**	**Max.**	**Nb. of obs.**
ADLs	0.248	0.857	0	6	181 620
IADLs	0.344	1.06	0	7	181 620
GALI	1.133	1.093	0	3	181 620
Mobility	1.633	2.349	0	10	181 333
Chronic diseases	1.718	1.55	0	14	181 307
Drinking	0.17	0.376	0	1	180 689

A.3 Technical progress

Table 7: Life expectancy at 65 years old for all waves and individuals (females and males)

Country	Mean	Std. dev.	Min.	Max.	Nb. of obs.
Austria	20.015	1.719	17.3	21.7	11 976
Germany	18.242	2.899	11.9	21.2	12 600
Sweden	19.728	1.48	17.4	21.3	12 149
Netherlands	19.289	1.776	16.3	21.2	12 306
Spain	20.841	2.211	17.2	23.4	14 297
Italy	20.333	1.956	17.3	22.6	13 140
France	21.286	2.301	17.7	23.8	15 869
Denmark	18.531	1.529	15.9	20.4	10 423
Greece	18.673	1.428	16.9	20.1	5 618
Switzerland	20.652	1.551	18.2	22.6	8 989
Belgium	19.469	1.884	16.5	21.6	17 449
Israel	20.256	1.111	18.7	21.3	4 671
Czech Republic	17.566	1.819	14.3	19.3	13 649
Poland	17.358	2.199	14.5	19.9	4 111
Luxembourg	20.581	1.398	19.8	21.9	1 590
Hungary	16.547	1.985	14.3	18.3	2 974
Portugal	19.918	1.888	17.8	21.6	1 920
Slovenia	19.422	2.089	16.9	21.4	5 525
Estonia	18.023	2.564	14.3	20.3	12 364
Total	19.431	2.335	11.9	23.8	180 620

A.4 Exogenous Shock

Table 8: Exogenous shock of income per country

Country	**Gift 250€ or more**			**Gift 5000€ or more**		
	Yes (%)	No (%)	Nb. of obs.	Yes (%)	No (%)	Nb. of obs.
Austria	10.12	89.88	8 726	13.55	86.45	8 699
Germany	7.69	92.31	8 510	21.98	78.02	8 485
Sweden	7.40	92.60	8 646	27.11	72.89	8 622
Netherlands	4.23	95.77	8 561	17.50	82.50	8 533
Spain	3.28	96.72	9 463	9.95	90.05	9 438
Italy	6.48	93.52	8 820	10.06	89.94	8 803
France	3.92	96.08	11 047	13.07	86.93	11 034
Denmark	8.45	81.55	7 282	25.59	74.41	7 262
Greece	11.03	88.97	3 990	19.98	82.02	3 960
Switzerland	5.77	94.23	6 487	23.23	76.77	6 475
Belgium	4.84	95.16	12 296	23.77	76.23	12 237
Israel	11.61	88.39	2 692	6.19	93.81	2 697
Czech Republic	0.01	99.99	9 310	11.78	88.22	9 295
Poland	7.42	92.58	2 805	9.81	90.19	2 802
Luxembourg	10.52	89.48	1 188	25.97	74.03	1 182
Hungary	5.35	94.65	1 963	15.32	84.68	1 952
Portugal	4.32	95.69	1 252	13.40	86.60	1 246
Slovenia	4.14	95.86	4 054	13.74	86.26	4 053
Estonia	9.40	90.60	8 391	7.08	92.92	8 395
Total	9 019	116 464	125 483	20 583	104 587	125 170

Table 9: Exogenous shock of income per wave

Waves	**Gift 250€ or more**			**Gift 5000€ or more**		
	Yes (%)	No (%)	Nb. of obs.	Yes (%)	No (%)	Nb. of obs.
Wave 1	5.43	94.57	20 016	28.53	71.47	19 905
Wave 2	7.10	92.90	23 400	14.89	85.11	23 378
Wave 4	7.20	92.80	38 745	16.30	83.70	38 651
Wave 5	8.04	91.96	43 322	11.85	88.15	43 236
Total	9 019	116 464	125 483	20 583	104 587	125 170

A.5 Instrumental variables for the income

Table 10: Unemployment rate

Waves	Wave 1		Wave 2				Wave 4			Wave 5
Country	2004	2005	2006	2007	2009	2010	2010	2011	2012	2013
Austria	5.49		5.25	4.86				4.6		5.3
Belgium	8.39	8.44	8.25	7.46				7.1		8.4
Czech Republic			7.15	5.32				6.7		7
Denmark	5.51		3.9	3.8				7.6		7
Estonia							16.7	12.3		8.6
France	8.47	8.49	8.45	7.66				8.8		9.9
Germany	9.79		10.25	8.66				5.8	5.38	5.2
Greece	10.59	9.99	9.01	8.4						
Hungary								11		
Israel					7.54	6.64				6.2
Italy	8		6.78	6.08				8.4		12.1
Luxembourg			4.73	4.07						5.8
Netherlands	4.56		3.91	3.18				5		7.2
Poland			13.85	9.61				9.6	10.09	
Portugal								12.7		
Slovenia								8.2		10.1
Spain	10.97		8.45	8.23				21.4		26.1
Sweden	6.53	7.48	7.07	6.16				7.8		8.1
Switzerland	4.3		4	3.6				4		4.4
Nb. of country	11		14				16			15

Table 11: Location of the main residence (%)

	Big City	**Suburbs**	**Large town**	**Small town**	**Rural area**
Wave 1	13.97	18.34	18.65	25.47	23.58
Wave 2	16.25	15.71	19.65	22.41	25.98
Wave 4	14.99	10.47	16.47	23.94	34.13
Wave 5	14.37	11.92	16.91	25.74	31.05

A.6 Instrumental variables - first step estimation

Table 12: First step estimation of the IV method

Variables	**Coefficients**
Dependant variable:	Log of income$_{t-1}$
Big city	0.072 ***
	(0.019)
Suburbs	0.241 ***
	(0.019)
Large town	0.132 ***
	(0.017)
Small town	0.018
	(0.015)
Rural area	*Reference*
Unemployment rate	-0.052 ***
	(0.001)
Nb.of obs.	57 847
Nb. of groups	41 407
Wald Chi2(5) - stat	1716.22
Prob > Chi2	0.000 ***

Taking jointly, all the coefficients are significant.