Network for Studies on Pensions, Aging and Retirement

Is Good Health Purchasable by Out-of-Pocket Money in Western Europe?

Evidence from SHARE

Hai Minh Tran

NETSPAR ACADEMIC SERIES

BSC Thesis 2015-051

Is good health purchasable by out-of-pocket money in Western Europe?

Evidence from SHARE

Hai Minh Tran

Supervised by
Robin L. Lumsdaine

A thesis submitted for the degree of
Bachelor of Econometrics and Operation Research

Erasmus Universiteit Rotterdam

July, 2015

Abstract

Good health is vital for all the living and there is an insatiable demand for it, especially among the elderly because health generally declines with age. This paper investigates whether people can buy higher quality medical care and in exchange, receive improvements in health perception. The analysis covers the trade of out-of-pocket (OOP) health expenditures for improved self-assessed health status by using two samples drawn from the data of the Survey of Health, Ageing and Retirement in Europe (SHARE) - which is representative for people at the age of 50 and older. Two models are introduced, where model 1 embraces a 2-part model and model 2 considers a tobit model; then, some econometrics issues are addressed to obtain more efficient estimates. As a result, sample 1 shows a positive but weak sign for the effect of OOP medical expense on health perception improvements, whereas in sample 2 a much stronger but not large in magnitude relation is found in the negative direction. That is in the latter sample, people are less likely to have improvements in health status when OOP expenses are increased so they "cannot buy good health with these payments". Robustness checks confirm these findings: they suggest the ability to purchase an improved health status by OOP expenditures in 2004/05 (sample 1) but not in 2006/07 (sample 2). However, empirical evidence show clearly that higher elasticities in the OOP medical expenditure in the years 2004/05 and 2006/07 corresponds to an improved health status by 2006/07 and 2010/11 in Greece for both sample 1 and sample 2 respectively. This could be explained by the fact that OOP expenses - in the form of informal payments (such as gratitude money) - in Western-Europe are anecdotal while in Greece it is widespread [Dixon et al., 2002, p. 23].

Acknowledgements

I would like to thank my supervisor, Professor Lumsdaine for her time and effort in helping me by her valuable comments to complete this thesis. Also, SHARE because: *This paper uses data from SHARE wave 4 release 1.1.1, as of March 28th 2013 (DOI: 10.6103/SHARE.w4.111) or SHARE waves 1 and 2 release 2.6.0, as of November 29th 2013 (DOIs: 10.6103/ SHARE.w1.260 and 10.6103/ SHARE.w2.260) or SHARELIFE release 1.0.0, as of November 24th 2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N 211909, SHARE-LEAP, N 227822 and SHARE M4, N 261982). Additional funding from 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 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see `www. share-project. org` for a full list of funding institutions).*

1 Introduction

Health is generally deemed to be crucially important, not only for personal interests but also for policy making, to operate a sound economy. The Constitution of WHO (1946) also claims that "good health is an essential component of development" and it is "vital to a nation's economic growth and internal stability".[1] Nevertheless a highly respected author, Anne Wilson Schaef, quoted that *"Good health is not something we can buy. However, it can be an extremely valuable savings account."*[2] [Lyseight, 2010, p.129]. The analysis of this thesis is restricted to the first part of the quote, to find empirical evidence for the validity of this statement.

Intuitively, if people (informally) pay general practitioners or physicians to increase quality in medical care, they might undergo better treatments, because patients are possibly taken more seriously. Moreover, those in the lower quantile of income or wealth can only afford lower quality of health care [Marshall et al., 2011], which might be associated with a lower self-perceived health status.

Hence, in most countries of Central and Eastern-Europe, informal payments are more popular [Szende and Culyer, 2006] to ensure this higher quality medical care, which as a result could signify an improved self-assessed health status. For example in Hungary [Szende and Culyer, 2006], Poland [Chawla et al., 1998], Bulgaria [Balabanova and McKee, 2002] or even in Greece [Calltorp and Abel-Smith, 1994]. Although in Western-Europe the 'evidence' for informal payments could be anecdotal [Dixon et al., 2002], there are signs for these payments to some extent such as in France [Bellanger and Mossé, 2000].

However, informal payment is not the only component of out-of-pocket (OOP) health expenditures but also inpatient-outpatient care, prescription drugs and etc.[3] And these OOP payments show to be positively correlated with the total health expenditures, that according to [WHO, 2003, p.111], is the highest in EU countries when comparing to CIS and CSEC[4]. Although there is not much evidence on sky-rocketing informal payments in Western Europe, in Germany for instance the amount of OOP medical expenses has risen steadily in the last years [Bremer, 2014].

In general, health declines with age and OOP medical expenditures can become reasonably high towards end of life; moreover in the period of 1998-2006, these amounts "are both pervasive and large" in the U.S. [Marshall et al., 2011]. Hence, the investigation of people with older age is somewhat more interesting because the gradient of these expenditures from one period to another might be immense, in comparison to younger people. Moreover, high OOP expenses on health can be an issue as it can put people at a great financial risk [You and Kobayashi, 2011]. In some countries, this matter is even a greater concern as individuals with lower income tend to pay proportionally more through these expenditures than those with higher income [Szende and Culyer, 2006].

Despite the disadvantages of the sizeable health expenditures, more OOP payments might well reflect the improvement of one's health status, especially in Central-Eastern-Europe (in the form of informal payments). Although, these health expenses are also significant in the Western part of Europe, little is known about the presence of higher quality medical care that arises from these OOP expenditures.

[1] See `http://www.who.int/trade/glossary/story046/en/`

[2] See `http://www.livinginprocess.com/anne-wilson-schaef.php` for more on Anne Wilson Schaef.

[3] The data provides information on these OOP expenditures: inpatient, outpatient care, prescription drugs, in day-care centers, nursing home and home-based care.

[4] CIS stands for Commonwealth of Independent States and CSEC covers 15 Central and South-Eastern EU countries.

The main aim of this thesis is to seek evidence on the effect of these direct payments in the improved health conditions. To put it differently, the research question of interest is: *To what extent does out-of-pocket medical expenditures play a role in the improvement of the self-assessed health status in Western Europe?*

A brief outline of this thesis as follows: *i*) first of all, the literature concerning this topic is discussed, then *ii*) the available data is presented with the design of the sample and the variables of interest, *iii*) then the paper proceeds with the methodology, where the model and the estimation method is elucidated along with some econometrics issues to consider, afterwards *iv*) the results and the findings are investigated, then *v*) some robustness and sensitivity analysis is performed on the baseline models, and finally *vi*) the conclusion with the discussion and further recommendations on this topic completes this thesis.

1.1 Literature review

There is voluminous amount of literature on health and OOP health expenditures because they are important topics for policy making strategies. The latter one is at special interest in the U.S., but it is becoming increasingly prominent in Europe as well (see [Bremer, 2014]).

However, very few discuss the relation between the two topics when the OOP payments are on the right-hand side (RHS) and health on the left-hand side (LHS). Moreover, even less consider the change of individuals' health status from one period to another in Western-Europe. Additionally, in literature mostly the amount and the negative consequences of these health expenses are researched, therefore, this thesis could make a valuable contribution towards this literature by analysing some of its positive effects.

First, the importance of health from various aspects is presented and how this topic fits in context. Nordhaus [2005] argues that the value of improvements in the health status of a nation's population is excluded from the conventional measures of economic performance, which shows how crucial health improvements can be. Jürges [2006] finds that the overall effect of health on wealth is positive but very small, whereas the relation of total OOP expenditures and wealth is much stronger [Marshall et al., 2011]. Johnston et al. [2009] states that there are significant differences between objective and subjective health measures, which could be useful for further analysis. Marmot [2005] finds evidence for gross inequalities in health between countries, which suggests the control of country indicators. Denton et al. [2004] discusses several models and determinants of health and affirm the importance of looking more closely at gender differences in health.

Next, You and Kobayashi [2011] and Bremer [2014] reveals several determinants of OOP expenditures in China and Germany respectively that allows to set up the generating process for this variable. Stewart [2004] finds evidence on the increase of long-term care expenditures with age and Marshall et al. [2011] concludes a significantly large total OOP spending towards end of life, particularly for higher age groups. In the latter, the authors used data from the Health and Retirement Study that is representative of the U.S. population age 50 or older. Sambamoorthi et al. [2003] states that in 1997, nearly 8% of the older population, more than 2.3 million people, spent greater than 10% of their income on prescription drugs. Szende and Culyer [2006]; Balabanova and McKee [2002]; Chawla et al. [1998] show significant private health expenditures through informal payments in Europe. Heijink et al. [2010] finds no evidence of the optimal survey design features in collecting data on OOP health expenses, which might question the validity of the information on this matter.

2 Data and Design

In this section, the analysis on the data and the sample are further elaborated in three subsections: section 2.1 describes the dataset itself and how the raw samples are constructed. Next, section 2.2 explains how and why the samples are further adjusted; these adjusted samples are the final and admissible samples that are used for the research. Lastly, section 2.3 discusses the main variable of interest and the key explanatory variables used for the research.

2.1 The dataset

The data at use is from the Survey of Health, Ageing and Retirement in Europe (SHARE), which is a longitudinal survey of more than 110,000 people from 20 European countries aged 50 or over; thus, this dataset is ideal for the analysis. SHARE has conducted 5 waves of interviews but wave 3 cannot be used for this research because it pays attention to people's life histories so it does contain all the relevant *key variables* as discussed later in section 2.3. In this way, the data consists of 4 waves, for which the interviews are made in 2004/05, 2006/07, 2010/11 and 2013.

	wave 1	wave 2	wave 4	**waves** 1&2	**waves** 2&4	All pooled
# Obs	30,817	36,731	59,599	47,099	77,199	85,841
Participated in all	30,817	36,731	59,599	20,449	19,131	11,130
		Countries where amounts are not denominated by euro				
Sweeden	3,053	2,745	2,122	2,010	1,691	1,269
Denmark	1,707	2,616	2,393	1,249	1,800	896
Switzerland	1,004	1,462	3,786	696	1,073	507
Israel	2,598	2,464	-	1,826	-	-
Czechia	-	2,830	6,196	-	1,377	-
Poland	-	2,468	1,880	-	1,647	-
Remainder	22,455	22,146	43,222	14,668	11,543	8,458
		Non-Western-European countries (with amounts denominated by euro in dataset)				
Hungary	-	-	3,076	-	-	-
Slovenia	-	-	2,756	-	-	-
Estonia	-	-	6,828	-	-	-
Remainder	22,455	22,146	30,562	14,668	11,543	8,458
After imputation	22,454	21,011	30,010	14,667	11,135	8,155
			A special case, Greece			
Greece	2,898	3,243	-	2,280	-	-
Western Europe	19,556	17,768	30,010	**12,387**	**11,135**	8,155

This table describes six stages (indicated by **bold** in the first column). First, the total number of observations (**# Obs**) is shown then respondents who **Participated in all** waves in case of a merger, are retained. Third, countries (in *italics*) where amounts are not denominated by euro are subtracted and **Remainder** shows the number after this deduction. Fourth, non-Western European countries are removed (in *italics*); fifth, unmatched observations are deleted **After imputation** files are merged and retained observations are shown. Finally observations for **Western-Europe** are presented in all cases.

Table 1: Summary of SHARE waves and raw sample construction

In order to consider the change in the state of health, at least two waves need to be used to construct one sample. Unfortunately, wave 4 does not contain a *key variable*, the OOP health expenses, hence, explanatory variables cannot be built from 2010/11. This means the data on the change in health perception from wave 4 to wave 5 cannot be used and thus wave 5 is not included in the analysis. As a result, the analysis is restricted to two (raw) samples.

The first sample (*sample 1*) is constructed by merging wave 1 and wave 2 (denoted by **wave 1&2** in table 1) and the second sample (*sample 2*) involves merging wave 2 and wave 4 (denoted by **wave 2&4** in table 1). Table 1 summarises the number of individuals after each merge and country selections ('All pooled' merges all waves for the analysis together). In this table, although *Greece* is considered to be a Western-European country and its amounts are denominated by euro, its (2,280) observations are subtracted (from *sample 1*) because SHARE has not conducted interviews in 2010/11 (wave 4); but this country is analysed separately in section 5.2 - hence the special case. Therefore, *sample 2* cannot include this country (as it involves the merging wave 2&4) and 2,280 individuals are dropped from the first sample, *sample 1*. Further, notice the relatively larger number of non-imputed observations for wave 4 ($30,562-30,010=552$); this is because, conversely to wave 1 and 2 the variables for non-responding partner is no longer imputed.

Note from table 1, imputations for item non-response (missing values) - that are typically encountered in large household surveys - are used because the number of individuals with all answered survey questions are very small; this can lead to a loss of valuable information and less efficient estimates [Christelis, 2011]. As mentioned, the analysis focuses mainly on Western-European countries, however, only those are considered whose amounts are denominated by euro. The latter prevents any bias from the conversion of currencies when determining the health expenditures. Additionally in table 1, although *Hungary*'s local currency is not the euro (its currency is the forint), SHARE denominates its amounts in euros.

Next, table 2 summarises the remaining Western-European countries with the euro currency in all waves (for the analysis). Each sample (**Sample 1** and **Sample 2**), is created by merging the waves above them as indicated by *italics* in the table. It can be seen from table 2, that *Ireland* is only included in wave 2 and *Portugal* appears only in wave 4, therefore, when merging the waves (to create samples) those are dropped out. All in all, 7 countries are taken into account to represent Western Europe (which are indicated by **bold** in the first row of table 2).

	IE	PT	**AT**	**BE**	**DE**	**FR**	**IT**	**NL**	**ES**	**Total**
Wave1	-	-	*1,594*	*3,827*	*3,008*	*3,193*	*2,559*	*2,979*	*2,396*	19,556
Wave2	*1,134*	-	*1,192*	*3,169*	*2,568*	*2,967*	*2,983*	*2,661*	*2,228*	17,768
Sample 1	-	-	1,119	2,808	1,544	1,998	1,766	1,777	1,375	**12,387**
Wave2	*1,134*	-	*1,192*	*3,169*	*2,568*	*2,967*	*2,983*	*2,661*	*2,228*	17,768
Wave4	-	*2,080*	*5,286*	*5,300*	*1,572*	*5,857*	*3,583*	*2,762*	*3,570*	30,010
Sample 2	-	-	662	2,143	1,372	1,888	2,006	1,638	1,426	**11,135**

The frst row indicates the 2-letter country codes: Ireland (IE), Portugal (PT), Austria (AT), Belgium (BE), Germany (DE), France (FR), Italy (IT), the Netherlands (NL) and Spain (ES).

Table 2: Number of observations per country in each raw sample

The resulting number of observations for the first raw sample (*sample 1*) and the second raw sample (*sample 2*) are $n_1^R = 12,387$ and $n_2^R = 11,135$ respectively, where the superscript R denotes 'raw'. In the next section, these constructed raw samples are adjusted by taking out some observations to achieve a less noisy data. The number of observations in these 'adjusted' samples are denoted by n_1^A and n_2^A and should be distinguished from those of 'raw' samples.

2.2 Sample adjustment

This section elucidates how the raw samples are adjusted to obtain a less noisy data. First, it can be seen in table 2 from section 2.1, that there are $n_1^R = 12,387$ and $n_2^R = 11,135$ observations in the raw samples. Now, the process of adjustment considers dropping observations consecutively from the raw sample that meet the following selection criteria:

a) missing information on the key variables, which includes 'missing' (M), 'refusal' (R), 'don't know' (D) and 'other' (O) answers. Plus, if SHARE has defined them as 'wrong or implausible' value (W); these for instance could arise from a division by zero when determining the body mass index (BMI);
b) missing information on the dependent variable;
c) and observations are omitted if respondents' age is below 50.

	Education			Retired		Insurance			Social Int.				Family Status				
	missing (M)	refusal (R)	don't k. (D)	refusal (R)	don't k. (D)	PHI (R)	PHI (D)	SHI (R)	Give H. (R)	Give H. (D)	Activity (R)	Activity (M)	Mstat (O)	Mstat (D)	Sibling (R)	Sibling (D)	**Total**
Sample 1	13	3	2	1	1	1	8	4	0	2	1	13	12	1	2	5	69
Sample 2	64	8	16	2	0	1	3	4	2	5	3	44	0	0	1	2	155

Table 3

	Age	Events	Behaviour				Health evaluations			Dependent comp.		
	Below 50	Hospital (D)	Phactiv (D)	BMI (W)	BMI (R)	BMI (M)	Chronic (R)	Chronic (D)	Chronic (M)	SAH (R)	SAH (D)	**Total**
Sample 1	586	4	1	11	15	1	0	7	12	0	2	639
Sample 2	337	1	5	0	18	15	1	3	13	9	4	406

Mstat and Phactiv represents 'marital status' and 'physical activity' respectively.

Table 4

	# Obs from Raw samples	# Omissions from table 3	# Omissions from table 4	# Obs from Adjusted samples
Sample 1	$n_1^R = 12,387$	69	639	$n_1^A = 11,679$
Sample 2	$n_2^R = 11,135$	155	406	$n_2^A = 10,574$

Table 5: The construction of the admissible number of observations in each sample

Nevertheless, in case the missing information on the variable at hand is relatively large, it is not omitted during the adjustment of the samples. To put it differently, if the number of missing information (that is, the number of (M), (R), (D), (O) or (W) values) in a variable is statistically significant on a 1% significance level then it is retained in the sample. This can be determined by performing a simple t-test with the null hypothesis that the number of missing information is equal to zero. Next, tables (3, 4, 5) summarise the above mentioned adjustment process. The first two tables (3, 4) count the total number of omitted observations for each sample, then table 5 shows the reduction of the raw samples after the adjustment.

Table 5 shows the transaction from the raw to the adjusted sample sizes. In the first sample subtracts $69 + 639 = 708$ observations, where 69 and 639 are the total individuals removed from the first (raw) sample from table 3 and 4 respectively. Similarly, the second sample omits $155+406 = 561$ respondents, which results in the final admissible number of observations ($n_1^A = 11,679$ for *sample 1* and $n_2^A = 10,574$ for *sample 2*).

The largest reduction from the original sample is made by removing people below the age of 50 as presented in table 4; they represent 4.7% and 3.0% of *sample 1 and 2* respectively. This omission is necessary because the data from SHARE is only representative for those who are 50 or above. Therefore, the retained percentages from the raw samples are 94.3% for the first and 95.0% for the second sample.[5] As mentioned before, omitting observations might lead to a loss of valuable information and less efficient estimators in the analysis, however as the reduction is not huge this step might be justified without its drawbacks.

2.3 The key variables

2.3.1 Dependent variable

The LHS variable is constructed from the self-assessed health (SAH) on the US scale, which is coded as 5 (*poor*), 4 (*fair*), 3 (*good*), 2 (*very good*) and 1 (*excellent*). To design the indication of a positive gradient of the health status with this measure in a sample, one has to consider the SAH in both the *initial* and the *terminal wave*.[6] Improvement occurs when individuals' self-perceived state of health changes from a lower category into a higher one. To illustrate, in *sample 1* if people's state of health is *fair* in the *initial wave* (in 2004/05) but by the *terminal wave* (in 2006/07) they rate their health as *good*, *very good* or *excellent* the sign for an improvement in health perception follows. This upgrade can be described by a dichotomous variable ($\Delta^+ SAH$), 1 if the improvement occurs and 0 otherwise.

	Raw Sample			**Adjusted Sample**		
	Improved	Missing values	# Obs	Improved	Missing values	# Obs
Sample 1	2,483 (20.1%)	57 (0.46%)	$n_1^R = 12,387$	2,345 (20.1%)	28 (0.24%)	$n_1^A = 11,679$
Sample 2	2,414 (21.7%)	51(0.46%)	$n_2^R = 11,135$	2,297 (21.7%)	27 (0.26%)	$n_2^A = 10,574$

This table shows the number of improvements in self-assessed health status (Improved), the missing values (M), and the number of observations (# Obs) for each sample, raw and adjusted. The values in brackets represents the percentages of the sample.

Table 6: Summary of the dependent variable in the raw and adjusted samples

Table 6 describes the dependent variables both for the first and second (raw and adjusted) samples. As mentioned before, the adjusted samples consist of less number of observations, hence less number of improvements, because some are dropped from the raw samples (see section 2.2). It can be seen in

[5] For instance for *sample 1* the calculation, as indicated in table 5 the raw sample is $n_1^R = 12,387$ and the new adjusted sample is $n_1^A = 11,679$, thus 94.3% is retained.

[6] *Initial waves* are from 2004/05 (wave 1) and 2006/07 (wave 2) and *terminal waves* are in the years 2006/07 (wave 2) and 2010/11 (wave 4) for sample 1 and 2 respectively.

table 6 that after the removal of some noise (particularly 708 and 561 observations are taken out of *sample 1* and *sample 2*), the percentages do not change at all for the number of upgrades (rounded off to two decimal places). However, the percentages for the number of missing values in each samples are reduced to almost their half. This suggests that the adjustment made on the samples is reasonable, since a less noisy data can be achieved. These missing values for each dependent variable are indicated by a binary variable on the RHS of the equation.

2.3.2 Independent variables

The key explanatory variables can be categorised into six major groups, depending on the part they explain. All the explanatory variables except for the OOP expenditure are categorical ones, which obtain a 1 when the individual belongs into that category and 0 otherwise. For instance, if respondents provide missing information and answered *"don't know"* or *refused* to answer or the value is *missing* for the variable *gender* then five different groups can be distinguished, hence, five binary variables are created.[7] The six groups with their subgroups that show importance in the explantation of health perception are explained in this section. The summary of these are presented in tables 11 and 12 in appendix A.

1) Demographics consist of ***i***) *Age and Gender* of the respondent. Since there is little difference in a one or two-year increment in age, 7 age brackets with five years of intervals until the age of 80 are constructed - the last group considers those who are 80 or older. *Gender* should be controlled in the analysis as Denton et al. [2004] find differences in health for male and female.

ii) *Location of individuals' main residence*, which can be in the city, the suburbs, in a small or large town or in a village could also play a major role in determining the atmosphere and the environment they live in, hence it could affect one's state of health.

iii) Note, Crystal et al. [2000] states that those with limited education are experiencing higher OOP costs; this motivates the creation of education categories. *Education* variant describes respondents' education level according to the ISCED code that ranges from 0 till 6, where 6 is being the highest.[8] This variant also contains information on those who are still in school, which is classified as lower than 0 ISCED code. This way three education level categories are distinguished: low, intermediate and high that corresponds to an ISCED code of 0-2 or lower, 3-4, and 5-6 respectively.

Career involves people's net income in previous year, which is divided into four quantiles with three cut points: the 25th, 50th and the 75th percentile because as Szende and Culyer [2006] claims OOP expenses are greater concern for those in the lower quantile of income. Further, their job situation is included (either retired or not retired) as people's current main activity is important in determining their health perception [Denton et al., 2004].

iv) Finally, *Insurance* status plays a great role in determining health expenditures. Intuitively, one would think that these expenses are the lowest when one has health insurance; like Crystal et al. [2000] argues, OOP healthcare cost burdens are the heaviest in the absence of these coverages. However, You and Kobayashi [2011] finds that there is not enough evidence to show that health insurance reduces OOP payments. The authors elucidate this finding by adverse selection and moral hazard. The former tells that people with high risk lifestyles are more likely to apply for an insurance program; whereas the latter concludes that insured individuals' levels of consciousness

[7] Because in addition to the 3 mentioned groups, gender can take two possibilities, male or female.

[8] ISCED stands for the international standard classification of education, for more information see: [Unesco, 1997].

in medical expenditures is lower (and thus they have greater willingness to pay for health) than those who have to finance all the charges for medical services by themselves [Rubin and Koelln, 1993]. Therefore two types of insurance coverages are distinguished in this analysis: private health insurance (PHI) - that also embraces voluntary and supplementary insurance - and statutory health insurance (SHI), which is called national health insurance as well.

The next part discusses the remaining 5 groups of the key variables.[9] As Denton et al. [2004] states, there are many determinants of health, such as physical, social or psychological ones, because health is a multi-dimensional concept. The authors also claim explanatory powers of measures for health related behaviours. Therefore, this analysis differentiates both social and psychological determinants for the improvements in health status and a lifestyle measure, that involves physical health outcomes, behaviours and other health measures.

2) Social determinants of health consist of two subgroups that describe the social surroundings of individuals. ***i*)** First, *Family* includes respondents' marital status, if they live alone (single) or together with someone (either a spouse or a partner). The number of children, grandchildren they have; these two variants are divided into three categories: if they have none, 1 or 2 and 3 or more (grand) children. Lastly, if they have any siblings or not. Generally, more family members could correspond to a improved health status.

***ii*)** Second, *Social interactions* tells if respondents have given or received any help (i.e. care, practical or administrative tasks) in the household (hhd) in the last one month.[10] Furthermore, the activity covariant indicates whether individuals have done any social activities outside of household in the last one month; this involves voluntary or charity work, if they have gone to sport, taken part in religious organisations and etc.

3) Psychological determinant of health, considers: ***i*)** *Make ends meet*, that reflects the financial distress of the respondents if any. An indication for a financial distress occurs when individuals make ends meet with some or great difficulty and consequently there is no sign for distress if they make ends meet fairly easily or easily.

4) Lifestyle measures, embrace indicators for ***i*)** *Health related behaviours*, which Denton et al. [2004] describes by the physical activity and the body mass index (BMI). The former shows no physical activity if respondents never or almost never engaged in neither moderate nor vigorous physical activity. The latter health related behaviour, the BMI is divided into four categories set by the World Health Organisation, to differentiate between the four body structures: underweight, normal, overweight or obese and for easier interpretations (see table 12 in appendix A for intervals).

***ii*)** *Health evaluations* encompass two self-reported health measures. The first one is represented by the number of chronic diseases, which includes heart attack, cancer, stroke etc. And the second one is the Euro-Depression (Euro-D) score, where Euro-D involves 12 indicators (the score equals to the number of items selected by the respondent.[11]).

***iii*)** Finally, *Recent health event* indicates if individuals had any serious health related problems that made them go to the hospital in the last 12 months.

[9] This is summarised by table 12 in appendix A with their descriptive statistics; the mean and the standard deviation (**Std.**), for the (adjusted) *sample 1* and *sample 2*

[10] This variant could be updated by the help given or received outside of the hhd, but SHARE did not collect information on this in later waves, only in wave 1.

[11] For more exact definition of Euro-D, see SHARE Release Guide 2.6.0 of waves 1&2 (`http://www.share-project.org/fileadmin/pdf_documentation/SHARE_guide_release_2-6-0.pdf`).

Further, Marmot [2005] finds gross inequalities in health between countries, therefore, binary indicators for the seven **5) Western European countries** are created to control for these differentials.

6) Other RHS variables, incorporate the ***i)*** *Missing values of SAH*, that are described in table 6 in the adjusted sample column. Although there are not many (0.24% and 0.26% in *sample 1* and *sample 2* respectively), they are included in the samples to achieve a less biased sample [Denton et al., 2004].

ii) *OOP health expenditure*, that is constructed by the sum of the expenditures of inpatient, outpatient care, prescription drugs, in day-care centres, home care and home-based care (not counting health insurance premiums) in the last 12 months. Because of the right skewed distribution of this variant, the log is taken, however, it should be noted that 0 values become minus infinity ones, that cannot be included in a regression analysis.[12] The remedy for this issue is outlined later in section 3.3. Additionally, this transformation allows for easier interpretation of this continuous variable.

During the constructions of the explanatory variables, some difficulties should be handled. The data on the income in *sample 1* is collected in gross amounts, while in the second sample the net amounts are provided. To solve this problem, the variables *"taken home before tax"* and *"taken home after tax"* are used to calculate the average tax rates in each country to obtain the net income in *sample 1*.[13]

In addition, the raw dataset provides the total income in the hhd, thus incomes from other hhd members should be subtracted to determine the individual earnings. Furthermore, some variables are not re-measured in later waves because those are not likely to change over that period. These variables are the height (for BMI) and the insurance status, hence the variables have to be incorporated from earlier waves.

3 Methods

This section elucidates the methods used for the analysis in three main stages. First, section 3.1 presents the models, then the second stage addresses some econometric issues (see section 3.2). Finally, the last stage presents the methodology for parameter estimations in section 3.3.

3.1 Model

To be able to present the baseline models for researching the effect of OOP expenditure on $\Delta^{+}SAH$, two models are introduced. First, *model 1* considers three equations while the the second model, *model 2* involves two. Section 3.1.1 describes the components of these considered models: the *main equation*, which models the relation of the *key variables* (this equation is present in both models), and the generating process of the OOP medical expenses. The latter one, is modelled by a 2-part model (*model 1*) and by a tobit model (*model 2*). Then section 3.1.2 relates the theoretical analysis with the problem at hand and completes the description of the models for this research.

[12] The "log", throughout this paper denotes the natural logarithm.

[13] The reason why country averages are used is because of the large scale of missing values for the variables *taken home before/after tax*.

3.1.1 Theoretical analysis

First, one important assumption has to be made; the indication for improvements in health status does not explain the continuous variable: the OOP payment. The following analysis derives why conjecture can be made. Consider Eq.(1) and (2), where a binary dependent variable depends on a continuous one and vice versa. This situation is more complex than the estimation of linear models with endogenous variables because a simple reduced form does not exist [Kim and Whasington, 2006]. This relation is described as follows:

$$y^*_{1i} = \beta_1 log(y_{2i}) + \sum_{k=1}^{m} \gamma_k x_{ki} + \eta_i \tag{1}$$

$$y_{2i} = exp\left(\beta_2 I_{y^*_{1i}>c_1} + \sum_{l=1}^{p} \gamma_l x_{li} + \epsilon_i\right) = exp\left(\beta_2 I_{y^*_{1i}>c_1} + \sum_{l=1}^{p} \gamma_l x_{li}\right) u_i \tag{2}$$

where the indicator function $I_{y^*_{1i}>c_1}$ obtains a 1 if the continuous latent variable y^*_{1i} is larger than a threshold c_1 (0 otherwise), y_{2i} is a continuous dependent variable and x_{ki} and x_{li} are m and p number of exogenous variables respectively. η_i and ϵ_i are disturbance terms with a normal distributions. Note, that η_i follows a standard normal, while the other error terms' parameters are not necessarily those of standard normal.[14]

In case both β_1 and β_2 are non-zero, then this system becomes a (full) simultaneous equation model (SEM), however, Kim and Whasington [2006] show that it cannot happen due to the logical consistency of Eq.(4), which is derived by substituting Eq.(2) into the following expression:

$$\begin{aligned} Pr(I_{y^*_{1i}>c_1} = 1) = Pr\left[\eta_i > -\left(\beta_1 log(y_{2i}) + \sum_{k=1}^{m} \gamma_k x_{ki}\right)\right] = \\ = \Phi\left[\beta_1 log(y_{2i}) + \sum_{k=1}^{m} \gamma_k x_{ki}\right] \end{aligned} \tag{3}$$

where Φ is the cumulative distribution function (CDF) of the standard normal. Then the resulting equality is combined with the property of the probability that:
$Pr[I_{y^*_{1i}>c_1} = 1] + Pr[I_{y^*_{1i}>c_1} = 0] = 1$, and the following condition can be derived by rewriting this property:

$$\begin{aligned} \Phi\left[\beta_1\left(\beta_2 * 1 + \sum_{l=1}^{p} \gamma_l x_{li} + \epsilon_i\right) + \sum_{k=1}^{m} \gamma_k x_{ki}\right] + \\ + \left\{1 - \Phi\left[\beta_1\left(\beta_2 * 0 + \sum_{l=1}^{p} \gamma_l x_{li} + \epsilon_i\right) + \sum_{k=1}^{m} \gamma_k x_{ki}\right]\right\} = 1 \end{aligned}$$

or:

$$\begin{aligned} \Phi\left[\beta_1\left(\beta_2 * 1 + \sum_{l=1}^{p} \gamma_l x_{li} + \epsilon_i\right) + \sum_{k=1}^{m} \gamma_k x_{ki}\right] = \\ = \Phi\left[\beta_1\left(\beta_2 * 0 + \sum_{l=1}^{p} \gamma_l x_{li} + \epsilon_i\right) + \sum_{k=1}^{m} \gamma_k x_{ki}\right]. \end{aligned} \tag{4}$$

[14] The standard normal distribution of η_i is needed for the identification of parameters in equation [Heij et al., 2004, p. 441].

Hence, fully SEM cannot be achieved because either β_1 or β_2 equals to 0 [Winkelman, 2003]. For this analysis, β_2 is assumed to be 0, so that the dichotomous variable $I_{y_{1i}^*>c_1}$ does not explain the continuous variable $log(y_{2i})$.[15] The goal of the research is to find an estimate for β_1 because that coefficient determines the scale of the effect of the OOP health expenditure on the improved health indicator ($\Delta^+ SAH$). This is elucidated in more detail next, in the section.

3.1.2 Empirical analysis

In literature as Nordhaus [2005] states, the value of improvements in the populations' health status is vital for measuring economic performance. Thus, to examine the relationship between the dependent variable $\Delta^+ SAH$ and the OOP payments, let $I_{y_{1i}^*>c_1} = \Delta^+ SAH$ the binary and $log(y_{2i})$ be the (continuous and normally distributed) log of OOP expenses from section 3.1.1 - see appendix C.

This way, the continuous and unobserved (to the researcher) variable y_{1i}^* can be associated to the change of the time spent outside of the household. For instance in *sample 2*: if people spend more time outside by 2010/11 to socialise or to do something instead of staying home, they could gain more pleasure or enjoyment compared to what they had in 2006/07. This could increase their self-assessed health status from a lower to a higher category. The measurement of this variant could be by questioning respondents on the time spent on certain activities (outside of household); although this could involve large measurement errors.

Also, from Eq.(1) let m be the number of independent variables from section 2.3.2 for the underlying variable of $I_{y_{1i}^*>c_1}$ that are enumerated in tables 11 and 12 in appendix A; and p be the number of RHS variables for the log of OOP medical expenses indicated by a † symbol in those tables.[16] Note that, due to the log transformation of OOP expenditure some information is lost because the implausible (minus infinity) values, as is explained in section 2.3.2. Therefore the generating process of the (log) OOP variant is delineated by a 2-part model, where the first part evaluates its participation (whether OOP payment occurred or not) and the second part analyses its levels by Eq.(2). The former part is outlined by the following relation:

$$\tilde{y}_{2i} = \sum_{l=1}^{p} \gamma_l x_{li} + \upsilon_i \quad \text{with} \quad \tilde{y}_{2i} = \begin{cases} 1 \text{ if } log(y_{2i}) > 0 \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

where the independent variables are those of Eq.(2) and the error term is υ_i. In summary, to answer the research question two models are considered. First, *model 1* is a 3-part model, which is described by Eq.(1) - also referred to as the *main equation* - and Eq.(2, 5). Note that this model embraces the 2-part model [Eq.(2, 5)] as described in literature [Seshamania and Gray, 2004; Dormont et al., 2006; Kim and Whasington, 2006]. Second, *model 2* is somewhat simpler than *model 1* as it excludes Eq.(5) and thus only involves two equations: Eq.(1) and Eq.(2). However, the latter equation's dependent variable and error term are different; instead the log of OOP medical expense, $log(y_{2i}^*)$ is used with $log(y_{2i}^*) = max\left[log(y_{2i}), 0\right]$ and the error term (ϵ_i) is multiplied by a scale parameter σ.

[15] In section 3.2.3 it can be seen that this assumption is somewhat weak for the first sample and suggest approaches that are closely related to instrumental variable estimation in SEM [Kim and Whasington, 2006].

[16] These RHS variables are chosen according to the paper of Bremer [2014] because the author investigated the determinants of OOP health expenditure by using data from SHARE. Note, due to multicollinearity 1 variable should be suppressed for each subgroup in the regression analysis.

For the rest of the analysis, the estimations up to the *main equation* is named as the *first level estimation*, and consequently the estimation of the remaining equation [Eq.(1)] is called the *second level estimation*. [17] It should be noted, that the main goal of the analysis is to estimate the parameter β_1 in Eq.(1), that is to determine the (partial) effect of the OOP medical expenses on the indicator of the improved self-assessed health status after controlling for other covariants. The estimation procedure of this model is elaborated in more detail in section 3.3.

3.2 Issues to consider

3.2.1 Heteroskedasticity

In panel data homoscedasticity is often rejected, which reduces the efficiency of the estimates. In order to test for this phenomena, the Breush-Pagan test can be performed. The test (with null hypothesis of homoskedasticity) involves a 2-stage regression procedure, where in the second stage the squared residuals (from 1st stage) are regressed on all the explanatory variables and their squares.

As a result, a strong indication for heteroskedasticity can be observed with p-values of 0.000 in both samples; this suggests the use of the (robust) Huber-White standard errors during the estimation process. However it should be noted that due to the trade off between efficiency and consistency and the presence of heteroskedasticity, maximum likelihood estimators in non-linear models might be inconsistent because these models become mis-specified [Greene, 2002, p.673]. Such model mis-specifications can be tested by out-of-sample forecasts and addressed by adjusting the likelihood function, which correctly incorporates the exact form of heteroskedasticity [Giles, 2013].

3.2.2 Selectivity bias

Selection bias can arise from the restriction of the sample at use. In this case, the sample is restricted to (7) Western European countries whose amounts are denominated by euro as mentioned in section 2.1. However, this is not the only restriction to the sample. Note that the log of OOP health expenditures is taken and since that variable only takes 0 or positive values, the transformation results in minus infinity values. In this way, the test for sample selection bias (suggested by J. J. Heckman) with the null hypothesis of no selectivity bias is given by a selection (of the dependent variable) and a regression equation.

Nonetheless, the test for sample bias should be performed in *two phases* because of the *two levels of estimations* described in section 3.1.2. The *first phase* involves the test of the bias that comes from the log transformation and the *second phase*, tests the bias that arises from the selection of Western European countries. In the *first phase* the selection occurs when $\tilde{y_{2i}} = 1$, which essentially tells whether an individual has OOP expenditures or not. Given this selection by Eq.(5), the regression equation, Eq.(2) can be estimated by OLS with White standard errors.

To illustrate, the test performs if $E\left[\epsilon_i | x_{li}, \tilde{y_{2i}} = 1\right] = \rho_{\epsilon_i, \upsilon_i} \sigma_{\epsilon_i} \lambda(X_{n\times p} \hat{\Gamma}_{p\times 1}) = 0$, where ϵ_i is the error term in Eq.(2), $\rho_{\epsilon_i,\upsilon_i}$ is the correlation of the disturbance terms from Eq.(2) and Eq.(5), and σ_{ϵ_i} is the standard deviation of ϵ_i and finally, $\lambda(X_{n\times p}\hat{\Gamma}_{p\times 1})$ is the Inverse Mills Ratio (IMR) evaluated

[17] The 2-part model (in *model 1*) is represented by Eq.(2, 5), thus the *first level estimation* for *model 1* involves the estimation of these two equations and the *second level estimation* is the parameter estimation of Eq.(1). And the *first level estimation* for *model 2* involves Eq.(2) and its *second level estimation* is that of *model 1*.

at the linear predictions of Eq.(5) in matrix notation. This conditional expectation can only be achieved when $\rho_{\epsilon_i,\upsilon_i} = 0$; given that this relation does not satisfy, the estimates of γ_l from Eq.(2) suffer from selection bias. Essentially, the second (OLS) regression in the Heckman test determines the scale of the omitted variable (the IMR) bias.

The *second phase* however, can only be carried out when the linear predictions of the OOP medical expenses $[\widehat{log(y_{2i})}]$ from Eq.(2) are obtained. That is, the test can be performed after the *first level estimation* so that there are no minus infinity values for $\widehat{log(y_{2i})}$ as they are in $log(y_{2i})$. The estimation procedure of these predictions is outlined in section 3.3. Before illustrations, consider the selection equation of the *second phase* sample bias test:

$$D_i = \beta_1 \hat{y}_{2i} + \sum_{k=1}^{m} \gamma_k x_{ki} + \psi_i \tag{6}$$

where D_i is the dummy variable, which indicates if individual i is selected ($D_i = 1$) or not selected ($D_i = 0$), and the independent variables are those of Eq.(1).[18] Afterwards to illustrate again, Heckman tests if $E\left[\eta_i | \widehat{log(y_{2i})}, x_{ki}, D_i = 1\right] = \rho_{\eta_i,\psi_i}\sigma_{\eta_i}\lambda(X_{n\times m}\hat{\Gamma}_{m\times 1}) = 0$ holds, where η_i and ψ_i are error terms from Eq.(1) and Eq.(6) respectively, σ_{η_i} is the standard deviation of the former one, and ρ_{η_i,ψ_i} is their correlation. Furthermore, $\lambda(X_{n\times m}\hat{\Gamma}_{m\times 1})$ denotes the IMR evaluated at the linear predictions of Eq.(6) in matrix notation. [19]

In the *first phase* selection test, both samples reject the null hypothesis of $\rho = 0$, with *p-values* of 0.009 and 0.000 respectively. Hence, it can be concluded that there is sample selection bias when performing the *first level estimation* due to the log transformation of OOP expenditures. However the *second phase* Heckman test, tells that there is not enough evidence to reject the null because the *p-values* are 0.144 and 0.828 for the two samples respectively. As a result, OLS can provide consistent estimates for the *second level estimation* if no other bias is present.

3.2.3 Endogeneity

Finally, a common econometric issue, especially in empirical studies, the endogeneity is investigated. Since the *main equation* Eq.(1) contains a financial variable: the log of OOP expenditures, the exogeneity of this variable is likely be rejected as these variables tend to have reverse causality on health and thus inconsistent parameters are estimated. To add, the other health measures (number of chronic diseases and Euro-D) should also be tested.

To analyse this phenomena the Durbin-Wu-Hausman (DWH) test (with null hypothesis of exogeneity) can be applied, which involves a 2-stage regression: *1)* regress possible endogenous variables on their explanatory variables; *2)* regress the *main equation* on all RHS variables and on the residuals obtained from the first stage. Since only the significance of the residuals' parameters are of interest, the test works well with binary dependent variables when robust standard errors are applied. Also, in step *2)* the coefficients of the regression should not be interpreted, only their signs.

[18] Selection happens when observation is in one of the 7 Western European countries.

[19] In sum, the *first phase* Heckman test is basically the *first level estimation* with the Inverse Mills Ratio (IMR) in Eq.(2); its selection equation Eq.(5) is estimated by a probit and its regression equation Eq.(2) is by OLS. As for the *second phase* test, it considers Eq.(6) for selection and Eq.(1) for regression equation. Since the latter equation (which constitutes a linear probability model) involves the estimation of the *main equation*, its estimated $\hat{\beta}_1$ is also presented in **Results** section.

In both steps OLS regression is performed with robust standard errors. As a result, other health variants shows no correlation with the error term, but the coefficients of the residuals of the OOP medical expenses in the second stage regression are significant for *sample 1* and insignificant for *sample 2* with p-values of 0.017 and 0.077 respectively on a 5% level.[20]

3.3 Estimation method

As discussed in section 3.2 there are a few difficulties that should be addressed to gain as much efficiency and consistency as possible. First, heteroskedasticity can be corrected by using robust (Huber-White) standard errors for each estimation.

Second, to correct for sample selection bias that arises from the log transformation, two models are delineated to carry out the analysis. The first model, (*Model 1*) uses a 2-part model for the *first level estimation* (as described in section 3.1) to encounter this issue; this is because in the paper of [Dormont et al., 2006] the authors stated that according to Monte Carlo studies the 2-part model proves to perform better than the Heckman model, even when the latter is the data generating process.[21]

The second model (*model 2*) considers a tobit regression on the *second level estimation*, that is the regression of Eq.(2). This is done by censoring all values that are less than zero (to be zeros) and then obtaining the linear predictions of Eq.(2), $\widehat{log(y_{2i})}$.

In this way, both models are able to eliminate the selection bias from the log transformation of the OOP spendings after the *first level estimation*. *Model 1* carries this out by performing a probit and an OLS estimation of Eq.(5) and Eq.(2) respectively in this order (with robust standard errors). Afterwards, the predictions of the (log) OOP health expenditures $\widehat{log(y_{2i})}$ can obtained from Eq.(2), which are later used for the *second level estimation* of the parameters in Eq.(1) by another probit regression.

The estimation of *model 2* is somewhat similar to that of *model 1* with the difference that it excludes Eq.(5) and instead of the OLS estimation of Eq.(2), it performs a tobit regression of that equation. Then the next step is exactly the same as for the first model, it obtains $\widehat{log(y_{2i})}$ from Eq.(2) and eventually, conditioning these predictions a probit regression of Eq.(1) is done.

Third, it is difficult to find a remedy for endogeneity in empirical research, on the one hand because instruments can easily be correlated with the disturbance term or even more complex relations could exist. On the other hand due to the limited availability of data, omitted variable bias, measurement errors or simultaneity can occur. Therefore, in the main analysis of this paper endogeneity is present and parameters are estimated inconsistently (in *sample 1* at least). Nonetheless, estimates could be efficient and reliable confidence intervals can be obtained with the help of the Huber-White standard errors.

[20] This questions the validity of the assumption from section 3.1.1, that β_2 in Eq.(2) is indeed zero (at least for the first sample, where exogeneity of the financial variable is rejected).

[21] The main difference between the two models is that the sample selection model contains the IMR in the second regression (in the regression equation), and thus as stated in [Dormont et al., 2006], the performance of this model crucially depends on the degree of collinearity between the RHS variables and the IMR [Leung and Yu, 1996].

To summarise, *model 1* considers estimation in three parts. In the first part a probit regression of Eq.(5) is performed to obtain the predicted probabilities $Pr[\tilde{y}_{2i} = 1]$. Then having these determined, the predicted binary variable $\hat{\tilde{y}}_{2i}$ can be formed by a rule of thumb depending on the fraction of successes: $\hat{\tilde{y}}_{2i} = 1$ if $Pr[\tilde{y}_{2i} = 1] > C$ and 0 otherwise, for $C = \#(\tilde{y}_{2i} = 1)/n$, where $\tilde{y}_{2i}$ is defined in Eq.(5) and $n \in \{n_1^A, n_2^A\}$ [Heij et al., 2004, p.453].

Afterwards in the second part, condition on $\hat{\tilde{y}}_{2i} = 1$, Eq.(2) is regressed by OLS to obtain $\widehat{log(y_{2i})}$. As a result, plausible values can be predicted $[\widehat{log(y_{2i})}]$ for the *second level estimation*, instead of the minus infinity ones, which are the consequences of the log transformation of the OOP health expenditures; plus more observations can be included than using the original variable. This is the *first level estimation* of *model 1* as shown in the table 7 below in the top left box.

Then given the linear predictions of $\widehat{log(y_{2i})}$ the final part performs another probit regression, on Eq.(1) to approximate the effect ($\hat{\beta}_1$) of the key variable on $\Delta^+ SAH$ (as is shown in the right box, under *Second level estimation*). Note, the estimated $\widehat{log(y_{2i})}$ from the *first level estimation* are used to obtain consistent estimates for β_1 and γ_k in the *second level estimation* [Windmeijer and Silva, 1997].

	First level estimation	*Second level estimation*
Model 1	$Pr[\tilde{y}_{2i} = 1] = \Phi\left(\sum_{l=1}^{p} \hat{\gamma}_l x_{li}\right)$ $E[log(y_{2i}) \mid x_{li}, \hat{\tilde{y}}_{2i} = 1] = \sum_{l=1}^{p} \hat{\gamma}_l x_{li}$	Estimation of the *main equation:* $Pr[\Delta^+ SAH = 1] = \Phi\left(\hat{\beta}_1 \widehat{log(y_{2i})} + \sum_{k=1}^{m} \hat{\gamma}_k x_{ki}\right)$
Model 2	$E[log(y_{2i}^*) \mid x_{li}] = \sum_{l=1}^{p} \hat{\gamma}_l x_{li}$ with $log(y_{2i}^*) = max\,[log(y_{2i}), 0]$	

Table 7: The summary of the models estimation process

As for *model 2* the procedure is almost identical, except that it has one less equation in the *first level estimation* as presented in the bottom left box of the table 7 on the next page (the procedure is identical for *second level estimation*). Thus, Eq.(2) is not estimated given $\hat{\tilde{y}}_{2i} = 1$ and the dependent variable $[log(y_{2i})]$ is censored from below at zero, that is: $log(y_{2i}^*) = 0$ if $log(y_{2i}) < 0$ and $log(y_{2i}^*) = log(y_{2i})$ otherwise (or: $log(y_{2i}^*) = max\,[log(y_{2i}), 0]$).

Note the tobit (as is the probit) model assumes the error term of Eq.(2) - with $log(y_{2i}^*)$ as dependent variable and $(\sigma\epsilon_i)$ as error term - to be standard normal and given values for this equation's explanatory variables (x_{li} for $l = 1, 2, \ldots, p.$) the distribution of the dependent variable $[log(y_{2i}^*)]$ is mixed continuous-discrete (see [Heij et al., 2004, p.491]). In such models (tobit and probit) the maximum likelihood estimation is performed.

4 Results

This section presents the estimation results of the effect of the log of OOP medical expenditures on the dependent variable (Δ^+SAH), which indicates an improved self-assessed health status. Note, the whole estimation results are elucidated and presented in appendix D.

Table 8 shows the probit estimates (Coef.) of this effect (on the *second level estimation*), that is denoted by $\hat{\beta}_1$ in section 3, when the 2-part model (*model 1*) and when the tobit model (*model 2*) is applied on the *first level estimation.* Also, the marginal effects (M.E.) and the odds ratios are presented along with their robust standard errors (S.E.). Furthermore, the last two columns of table 8 also shows the results from the Heckman selection test on the *second phase.*[22] The table shows the estimates of the regression equation of the test, which is an OLS estimate with robust standard errors.

	2-part model estimates						Heckman	
model 1	Coef.	S.E.	M.E.	S.E.	Odds	S.E.	Coef.	S.E.
Sample 1	0.008	0.095	0.002	0.027	1.01	0.169	0.025	0.028
Sample 2	-0.150***	0.057	-0.044***	0.017	0.775***	0.076	-0.042*	0.025

	Tobit estimates						Heckman	
model 2	Coef.	S.E.	M.E.	S.E.	Odds	S.E.	Coef.	S.E.
Sample 1	0.001	0.013	0.000	0.004	1.00	0.022	0.004	0.004
Sample 2	-0.030***	0.011	-0.009***	0.003	0.951***	0.019	-0.007*	0.004

In all tables *, **, or *** indicates whether it is significant on 10%, 5% or 1% of significance level according to its p-value plus, all S.E. are that of the Huber-White (robust) standard errors with a rounding to three significant digits. Also, in later sections, these outcomes are referred to as the results of the original setting.

Table 8: Estimation results performed by the 3-part baseline model

When considering the first sample (*sample 1*), both *model 1* and *model 2* show somewhat positive but not significant effects for the OOP medical expense on Δ^+SAH in the probit regressions. The last two columns presents the results of the Heckman regressions, which share the same conclusion. This signifies that there is no evidence for the sign of the (partial) effect $\hat{\beta}_1$ after controlling other covariants. However, it should be noted from section 3.2.3 that the coefficients estimated in the first sample suffer from endogeneity, therefore they are more likely to be inconsistent.

The analysis on the second sample shows that, an increase in the elasticity of the OOP payments would not correspond to an improvement in health perception on 1% significance level for both models. The odds of being in an improved state of health versus not is 0.775 and 0.951 for the first (*model 1*) and for the second model (*model 2*). This essentially means that people are more likely to perceive themselves being in the same or worse health status in 2010/11 (in comparison to 2006/07) when the elasticity of OOP medical expenses increases.

[22] As discussed in 3.2.2, the *second phase* sample selection test is the one for testing the selection of Western-European countries, whose amounts are denominated in euros. Again note, the regression equation of this phase of the test, which is estimated by OLS, constitutes a LPM with the IMR. It should also be noted, that the error terms in the LPM are not normally distributed [Heij et al., 2004].

The marginal effects of the financial variable measures its instantaneous rate of change, which is -4.4% and -0.9% for *model 1* and *model 2* respectively. Although their directions are similar, the marginal effect of *model 1* does not fall into that of *model 2*'s 95% confidence interval (the lower bound is -1.5% and the upper bound is -0.2% so -4.4% is out of range).

Conversely, due to the larger standard errors in the 2-part model set up, this result of *model 2* falls into *model 1*'s 95% confidence interval. In the Heckman regressions, which corresponds to the estimates of a linear probability model (LPM) with the IMR, the marginal effects in both models (-4.2% in *model 1* and -0.7% for *model 2*) are not significantly different than those in the probit regressions (-4.4% and -0.9% respectively).

In sum, in the second sample both models show the negative direction of the researched effect, but, the magnitude is somewhat different. *Model 2* indicates a lower scale of the parameter than that of *model 1*. Since the standard errors in the second model is also lower it might suggest that the model is more robust and more efficient. Therefore, the effect of OOP health expenditures on the improvements of self-assessed health status are negative and not large.

5 Robustness and Sensitivity Analysis

5.1 A different self-assessed health measure

Here, the definition of the dependent variable is reconsidered to support the findings. Instead of comparing the two self-assed health statuses in the *initial* and the *terminal wave* in a sample, a variable in the *terminal wave* is considered. This variant measures how respondents perceive their health in comparison to last interview, with three entries: 'worse', 'about the same' and 'better'. This means the dummy for the improved health is generated by setting 1's for those who claimed a 'better' health and 0 otherwise. Unfortunately, this variable is only available in wave 2, hence only *sample 1* can be investigated.

	Probit regression						Heckman		
	Coef.	S.E.	M.E.	S.E.	Odds	S.E.	Coef.	S.E.	Improved
Sample 1	-0.134	0.111	-0.013	0.011	0.745	0.172	-0.010	0.016	611

Table 9: Results of 3-part model with different health measure

Table 9 shows the improved health indicator (last column) with only 611 positive indications, which is much less than the 2,345 (from table 6 in section 2.3.1). Also, note that the number of observations are reduced by 1, that is $n_1^A = 11,678$ because there is 1 'don't know' answer for this dependent variable (which is omitted from sample). Results of estimation are showed in table 9, where none of the values are significantly different than zero as it is with the original dependent variable $\Delta^+ SAH$. However, the direction of the researched effect becomes negative as in *sample 2* with the original setting (for more robustness checks see appendix B).

5.2 A special case: Greece

Finally, this analysis focuses on another Western-European country: Greece.[23] There are two reasons why this country is investigated separately. On the one hand, according to Calltorp and Abel-Smith [1994], in Greece OOP health expenditures are dominant especially in the form of informal payments. On the other hand, this is the only Western-European country (with the euro denomination of the amounts) that is left out of the first sample because *sample 2* does contain it.

The reason for this dissimilarity is that in wave 4 (in 2010/11) there are no interviews in Greece and consequently when merging wave 2&4 (to construct *sample 2*) the country drops out. Therefore, only *sample 1* has information on this and the resulting number of observations is $n_1^A = \mathbf{2{,}054}$ with **416** number of indications for the improvement of self-assessed health status.[24]

It should be tested for sample selection bias again (on the *second phase*), so the definition of D_i in section 3.2.2 changes to the selection of those who live in Greece (instead of the selection of people who live in the 7 Western-European country).[25] The test rejects sample bias as before with a p-value of 0.280.

Table 10 shows the results for both models. It can be seen that, the results of the two models coincide and they are significant at the 5% significance level. Unlike before in table 8 for *sample 1*, the results are all insignificant. This could be explained by the increased robust standard errors in comparison to the one presented in the original setting; clearly, this is because the substantial reduction of the sample size.

	Probit regression						Heckman	
1st level est.	Coef.	S.E.	M.E.	S.E.	Odds	S.E.	Coef.	S.E.
2-part model	0.884**	0.375	0.242**	0.103	4.370**	2.820	0.257**	0.120
Tobit model	1.270**	0.536	0.347**	0.147	8.240**	7.620	0.367**	0.172

Table 10: Results with 2 different *first level estimation* procedures

In comparison to the other 7 Western-European countries in the original setting, the effect of OOP health expenditure in Greece on the improved health is positive in direction and also larger in magnitude. This could be explained because OOP expenses, in the form of informal payments, in Western-Europe are anecdotal while in Greece it is widespread [Dixon et al., 2002, p.23].

[23] Although Greece is on the Easter part of Europe it is considered to be a Western-European country.

[24] Note in table 1 there are 2,080 observations that are subtracted from *sample 1*, but only 2,054 is used due to the adjustments in section 2.2.

[25] On the *first phase* the test rejects the null hypothesis but the two defined models can encounter this sample selection bias.

6 Conclusion

To summarise, in answer to the research question - *To what extent does out-of-pocket medical expenditures play a role in the improvement of the self-assessed health status in Western Europe?* - this thesis uses a 3-part model (*model 1*) and a model with tobit estimates on the *first level estimation* (*model 2*). These models are applied on two samples, which are cleaned and constructed from the data of SHARE to examine the relation (β_1) of OOP medical expenditures on the indicator for improved self-assessed health status ($\Delta^+ SAH$). First, the *first level estimation* has to be carried out, which *model 1* performs by a probit estimation of Eq.(5) and then an OLS regression of Eq.(2) with Huber-White standard errors. Whereas *model 2* performs a tobit regression on the modified version of Eq.(2). Second, having these parameters estimated, linear predictions for the OOP payments can be made (denoted as $\widehat{log(y_{2i})}$) to perform the *second level estimation*, which embraces (another) probit estimation of the *main equation*, to determine the desired effects (of the OOP medical expenses on $\Delta^+ SAH$).

Next, three econometric issues have to be addressed. There is a strong indication for heteroskedasticity, which more or less can be corrected by robust standard errors. Then, in the first sample the OOP payment is tested to be endogenous as it is a financial variable, which tend to have reverse causality on health variants. In this paper, this issue is not handled due to its difficulty of finding a remedy in empirical research. Last but not least, sample selection bias has to be tested. This is carried out in two phases, where the *first phase* embraces the selection of those who have OOP health expenditure and the *second one* considers the selection of the 7 Western European countries in table 2. The *first phase* showed selectivity bias, whereas in the *second phase* there is not enough evidence for the rejection of the null hypothesis.

As a result, the two samples not only show differences in the endogeneity test, but also in the outcomes of the estimations. In *sample 1* the effect of the OOP health expense is positive but weak, not significantly different than zero.[26] However, in *sample 2* this variable has much stronger signs in the opposite direction with a marginal effects of -4.4% and -0.9% and with the odds of being in the non-improved state of health versus being in one in the *terminal wave* are 0.76 and 0.95 (for *model 1* and *2* respectively).

Model 2 seems to be more robust and efficient as its standard errors are smaller, hence as it suggests these effects are not strong. The results also coincide when a different self-assessed health measure is considered. In addition, both samples show a strong empirical evidence on the positive researched effect in Greece. In other words, higher elasticities in the OOP medical expenditure in 2004/05 and 2006/07 corresponds to an improved health status by 2006/07 and 2010/11 for *sample 1* and *sample 2* respectively. This could be elucidated by the more popular OOP expenses (in the form of informal payments) in Greece (than in other Western-European countries) [Dixon et al., 2002, p.23].

To conclude, the validity of the quote from Anne W. Schaef can be questioned from the perspective of Greece because there are strong signs for achieving an improved self-assessed health by increasing the OOP medical expenditures. However, in other Western-European countries this does not hold as results suggest, only the same or worse health status can be reached, which could enforce the highly respected author's right.

[26] Note in the first sample the financial variable rejected the exogeneity property, thus parameters are inconsistent. Further, it considers the change of health from *initial wave* in 2004/05 to the *terminal wave* in 2006/07 while the second sample, the change from 2006/07 to 2010/11.

6.1 Discussion and Further Improvements

As pointed out in the data section, imputed values are used in this investigation to reduce the loss of valuable information and to obtain more efficient estimates [Christelis, 2011]. Another robustness check is done by leaving these values out and observe the change in the nature of the OOP health expenditures' effects (see appendix B). This analysis confirms the strong difference in the two samples; which in *sample 1* it tells that OOP medical expenses (in 2004/05) do have a positive contribution to health perception improvements, while in *sample 2* it suggest a clear (at least for *model 2*) negative effect (in 2006/07). That is the second sample rejects the possibility of achieving improved health status by increasing OOP health expenditure.

Nonetheless, there are a few drawbacks in the methodology. Although many researchers apply probit/logit maximum likelihood estimators, these estimators are biased in an unknown direction in the presence of any form of heteroskedasticity [Greene, 2002, p.673] because these models could become mis-specified. There are not many tests to diagnose this, thus out-of-sample forecasts could be considered.

Next, as discussed in section 3.2.3 that the assumption made in section 3.1.1 that the self-assessed health indicator variable $\Delta^{+}SAH$ does not explain the OOP health expenditure, is somewhat weak (for *sample 1* at least) and this suggests other approach that is closely related to instrumental variable estimation in SEM as is applied by Kim and Whasington [2006]. To remedy this weak assumption, the authors suggest applying limited information maximum likelihood estimators, and in another paper [Windmeijer and Silva, 1997] GMM estimation is used.

Although the solution for endogeneity could be the use of instrumental variables (IV), it might not be optimal to apply because instruments could also be correlated with other omitted variables that are not observed (by the researcher), which leads to difficulties in empirical research. However, these results could be compared with those with an instrument for the OOP payments (after testing its validity with the Sargan test) for further robustness check.

Also, some other improvements could be done on this topic to obtain more insights. In this thesis, self-perceived health measure is used, which could be altered by considering other health measures. As mentioned by Johnston et al. [2009], there are significant differences between objective and subjective (self-reported) health measures, which could also be included in a further analysis on this topic. However, the inclusion of more improved health indicators could suggest the use of SEM as these variables tend to be correlated with each other. The consideration of such models is generally more complex as multivariate normal distributions are evaluated and the problem of identification should be handled (see [Wissen and Golob, 1990] for the analysis of multiple equations).

Furthermore, multinomial or conditional logit/probit models could also be very prominent (instead of binary response variables) in this analysis because this way, more information on the outcome can be derived. Instead of two categories (whether to have improved health status or not), 3 distinct categories can be observed: 'better', 'worse' or 'about the same' health status. Hence, as *sample 2* shows a negative researched effect, which could be either 'worse' or 'about the same' health perception in this thesis, these suggested models could solve this dilemma.

Finally, the use of interaction terms in the regression analysis - to differentiate the effect of each country, gender or age group - could be very useful as well for this research as OOP health expenditures tend to be fluctuating for these different groups (see section 1.1).

Appendix

A Analysis with Descriptive Statistics

Table 11 summarises all the key explanatory variables discussed in section 2.3.2 and also includes the descriptive their statistics namely the mean and the standard deviation (**Std.**) in the adjusted samples. It can be seen in the table, that those statistics coincide well across the two samples, which could suggest non-conflicting results in the analysis between them. However, as the **Results** section shows, only the coefficients in *sample 2* are significant (for both models).

Number of observations: $n_1^A = 11,679$**,** $n_2^A = 10,574$

		Descriptive Statistics			
1) Demographics	**Description**	**Sapmle 1**		**Sample 2**	
i) *Age & gender* †		**Mean**	**Std.**	**Mean**	**Std.**
age 50/55	*1 if* $50 \leq age < 55$	0.19	0.40	0.17	0.37
age 55/60	*if it's between 55 and 60*	0.20	0.40	0.20	0.40
age 60/65	*if it's between 60 and 65*	0.18	0.38	0.18	0.38
age 65/70	*if it's between 65 and 70*	0.16	0.36	0.17	0.37
age 70/75	*if it's between 70 and 75*	0.12	0.33	0.13	0.34
age 75/80	*if it's between 75 and 80*	0.09	0.28	0.09	0.29
age 80+	*if* $age \geq 80$ *years*	0.07	0.25	0.06	0.24
male	*1 if respondent is male*	0.45	0.50	0.45	0.50
female	*if respondent is female*	0.55	0.50	0.55	0.50
ii) *Location*					
city	*1 if location is in the city*	0.11	0.31	0.11	0.31
large town	*if it's in a large town*	0.15	0.36	0.15	0.36
small town	*if it's in a small town*	0.30	0.46	0.30	0.46
suburbs	*if it's in the suburbs*	0.18	0.39	0.16	0.37
village	*1 if location is in a village*	0.26	0.44	0.28	0.45
iii) *Education & career* †					
low education	*1 if ISCED code* ≤ 2	0.53	0.50	0.52	0.50
intermediate education	*if code is 3 or 4*	0.28	0.45	0.29	0.45
high education	*if code is 5 or 6*	0.18	0.39	0.19	0.39
other education	*resp. select option: other*	0.01	0.07	0.01	0.07
income Q1	*lower than 25th percentile*	0.23	0.42	0.25	0.43
income Q2	*between 25th and 50th p.*	0.25	0.43	0.25	0.44
income Q3	*between 50th and 75th p.*	0.26	0.44	0.25	0.43
income Q4	*higher than the 75th p.*	0.26	0.44	0.24	0.43
retired	*1 if respondent is retired*	0.51	0.50	0.50	0.50
not retired	*1 if they are not retired*	0.49	0.50	0.50	0.50
iv) *Insurance* †					
SHI	*Statutory h. insurance*	0.76	0.43	0.56	0.50
no SHI	*1 if they do not have SHI*	0.09	0.28	0.07	0.25
SHI (D)	*don't know for SHI*	0.01	0.10	0.01	0.07
SHI (M)	*missing value for SHI*	0.14	0.35	0.37	0.48
PHI	*Private health insurance*	0.50	0.50	0.38	0.49
no PHI	*1 if they do not have PHI*	0.48	0.50	0.35	0.48
PHI (M)	*missing value for PHI*	0.02	0.14	0.27	0.44

In this table and table 12 the symbol † indicates the explanatory variables of the OOP health expenditure analogous to the paper of [Bremer, 2014], where the author also used SHARE data (discussed later for the analysis in section 3.1.2). The choice for the categories of education levels is also made according to this paper. Next, Q1 is the abbreviation for the first quantile, (D) stands for *'don't know'* answers and (M) for *missing* values.

Table 11: Demographic variables and their descriptive statistics

This table shows that more respondents are in the early ages of the second half of human life (from age 50 till age 70), more of them live in a small town or village (30% and 26% respectively) and that, more than 53% of them have lower education level, which is a considerable amount.

Furthermore, in the fourth subgroup, the insurance status show some differences. The means for SHI and PHI in the second sample (0.56 and 0.38) are considerably smaller than those of the first sample (0.76 and 0.50). The reason for this difference is that in wave 2 and in later waves, the existence of insurance coverages is not asked again but changes if any. In other words, in these waves respondents are asked if their coverage got 'worse', 'better' or stayed the 'same', hence, newcomers in those years are not asked about the insurance status. The missing (M) values for the two types of insurance coverages are significantly larger. In *sample 1* the means of SHI and PHI are 0.14 and 0.02, whereas in *sample 2* they are 0.37 and 0.27 respectively. This increase of missing values might explain some of the differences in the results of the research.[27]

Next, the three determinants of health that are discussed in section 2.3.2 are summarised in table 12, along with statistics (the mean and the standard deviation) for the two samples. It can be observed that the provided statistics match very well and that in the *Family* group, respondents are more likely to be surrounded by family members. In the first sample (*sample 1*), 75% of individuals live together with someone, 90% of them have siblings, 89% and 67% of them have children and grandchildren respectively. However in the same sample, only 6% and 4% of respondents have given and received help in the household respectively, which is very modest. This is due to the large scale of missing information on these variables (19% for help given and 62% for help received in *sample 1*). Note, due to the strong similarities in the two samples, the analysis for *sample 2* roughly coincides with that of *sample 1*.

Further in table 12, individuals in both samples seem to be in a relatively good health, that is (when considering the first sample again) 90% are physically active in some form, 20% (1% underweight and 19% obese) show abnormal body structure and 73% and 86% of respondents have no depression have not been in hospital for the past 1 year. Although, 43% of them have 2 or more chronic diseases, which is fairly considerable.

The statistics of the country indicators also match closely, although in *sample 1* there are slightly more individuals than in *sample 2* (with means of 0.23 and 0.19), and conversely, slightly more respondents in Italy in *sample 2* than in the first sample (with means of 0.14 and 0.18).

The means in the dependent variable's missing value indicator (on the RHS) show to have 0 values when rounded off to two decimals. Nevertheless, as table 6 in section 2.3.1 highlights, they represent 0.24% and 0.26% in the first and second sample respectively. Another key explanatory variable, the OOP health expenditure shows its means (that do not represents percentages anymore because this variant is continuous) of 5.04 and 4.98 for *sample 1* and *sample 2* respectively; this corresponds to the means of €154.5 and €145.5 in levels amount.

[27]Note, the means for each binary variables represent the percentages of 1's in the sample. This is not the case with OOP variable in table 12 because it is a continuous variant.

Number of observations: $n_1^A = 11,679$, $n_2^A = 10,574$

		Descriptive statistics			
2) Social Determinants	**Description**	**Sample 1**		**Sample 2**	
i) *Family*		**Mean**	**Std.**	**Mean**	**Std.**
alone/single	*single/lives alone*	0.25	0.43	0.23	0.42
together	*live with spouse/partner*	0.75	0.43	0.77	0.42
sibling	*1 if ever had any siblings*	0.90	0.31	0.90	0.29
no sibling	*don't have any siblings*	0.10	0.31	0.10	0.29
child 0	*not any children*	0.11	0.31	0.10	0.30
child 1/2	*1 or 2 children*	0.54	0.50	0.55	0.50
child 3+	*3 or more children*	0.35	0.48	0.35	0.48
gChild 0	*0 grandchildren*	0.33	0.47	0.34	0.47
gcChild 1/2	*1 or 2 grandchildren*	0.28	0.45	0.28	0.45
gChild 3+	*3 or more grandchildren*	0.39	0.49	0.39	0.49
ii) *Social Interactions*					
activity	*any activities outside hhd*	0.49	0.50	0.49	0.50
no activity	*not activities at all*	0.50	0.50	0.50	0.50
activity (D)	*don't know answer*	0.01	0.07	0.01	0.07
help given	*gave any help in hhd*	0.06	0.23	0.06	0.24
no help given	*not any help given*	0.75	0.43	0.76	0.43
help given (M)	*missing value*	0.19	0.39	0.18	0.38
help received	received any help in hhd	0.04	0.19	0.04	0.19
no help received	*no help received in hhd*	0.34	0.47	0.32	0.47
help received (M)	*missing value*	0.62	0.49	0.64	0.48
3) Psychological health determinant					
i) *Make ends meet* †					
no fdistress	*no financial distress*	0.64	0.48	0.64	0.48
fdistress	*resp. have financial distress*	0.36	0.48	0.36	0.48
4) Lifestyle measures					
i) *Health related behaviours* †					
no phactive	*no physical activity at all*	0.10	0.30	0.10	0.30
phactive	*some physical activity*	0.90	0.30	0.90	0.30
underweight	*BMI* <18.5	0.01	0.10	0.01	0.10
normal	$18.5 \leq$ *BMI* <25	0.37	0.48	0.36	0.48
overweight	$25 \leq$ *BMI* <30	0.42	0.49	0.42	0.49
obese	*BMI* >30	0.19	0.39	0.19	0.39
BMI (D)	*don't know value for BMI*	0.01	0.11	0.01	0.12
ii) *Health evaluations* †					
chronic <2	*0 or 1 chronic disease*	0.57	0.50	0.57	0.50
chronic ≥ 2	*2 or more chronic diseases*	0.43	0.50	0.43	0.50
no depression	*0-3 EURO-D score*	0.73	0.45	0.74	0.44
depression	*4 or more EURO-D score*	0.26	0.44	0.25	0.43
iii) *Recent health event* †					
hospital	*been in hospital last 12 months*	0.14	0.34	0.14	0.35
no hospital	*no hospital visit last 12 months*	0.86	0.34	0.86	0.35
5) Western European Countries †					
country (AT)	*Austria*	0.09	0.29	0.06	0.24
country (BE)	*Belgium*	0.23	0.42	0.19	0.40
country (DE)	*Germany*	0.13	0.33	0.12	0.33
country (ES)	*Spain*	0.11	0.31	0.13	0.33
country (FR)	*France*	0.16	0.36	0.16	0.37
country (IT)	*Italy*	0.14	0.35	0.18	0.39
country (NL)	*Netherlands*	0.14	0.35	0.15	0.35
6) Other RHS variables					
i) SAH (M)	*missing values of SAH*	0.00	0.05	0.00	0.05
ii) OOP	*log of OOP health expenditures*	5.04	1.46	4.98	1.41

Table 12: Other independent variables and the percentages of improved health

The following part, discusses tables 13 and 14, that again represents all the independent variables but now they emphasise the percentage of improved health (at a given variable) to the total number of observations for both (adjusted) samples. To illustrate, consider the male variant in table 13 for the first sample. There are 1,029 men who showed an improvement in health status, thus the percentage to the total is $(1,029/n_1^A) \times 100 = 8.8\%$ with $n_1^A = 11,679$ from table 5. Note that, these percentages of these explanatory variables do not match closely as their means, because number of observations are different and the number of improvements across samples are also different (although these are not major differences). Consider the same male variant in table 13 but now for the second sample. The provided percentage is 9.9%, which is considerably larger than that of *sample 1* (8.8%). Although, there are just slightly more men who have improved health status in *sample 2* (1,045), the division yields: $(1,029/n_2^A) \times 100 = 9.9\%$ with $n_2^A = 10,574$.

	Percentage of improved health to total observations		
1) Demographics	**Description**	**Sample 1**	**Sample 2**
Age & gender †			
age 50/55	*1 if age is between 50 and 55*	4.0	3.7
age 55/60	*if it's between 55 and 60*	4.2	4.8
age 60/65	*if it's between 60 and 65*	3.7	4.0
age 65/70	*if it's between 65 and 70*	3.0	3.6
age 70/75	*if it's between 70 and 75*	2.0	2.5
age 75/80	*if it's between 75 and 80*	1.8	1.9
age 80+	*if age is more than 80 years*	1.3	1.1
male	*1 if respondent is male*	8.8	9.9
female	*if respondent is female*	11.3	11.8
Location of main residence			
city	*1 if location is in the city*	2.0	2.2
large town	*if it's in a large town*	3.1	3.2
small town	*if it's in a small town*	6.6	6.6
suburbs	*if it's in the suburbs*	3.4	3.7
village	*1 if location is in a village*	5.0	6.0
Education & career †			
low education	*1 if ISCED code is 0-2 or lower*[28]	10.8	11.4
intermediate education	*if code is 3 or 4*	5.6	6.3
high education	*if code is 5 or 6*	3.6	3.8
other education	*resp. select option: other*	0.1	0.1
income Q1	*lower than 25th percentile*	4.6	5.5
income Q2	*between 25th and 50th p.*	5.0	5.5
income Q3	*between 50th and 75th p.*	5.3	5.7
income Q4	*higher than the 75th p.*	5.1	5.0
retired	*1 if respondent is retired*	9.7	10.3
not retired	*1 if they are not retired*	10.3	11.4
Insurance †			
SHI	*Statutory health insurance*	15.3	11.8
no SHI	*1 if they do not have SHI*	1.7	1.6
SHI (D)	*don't know answer for SHI*	0.2	0.1
SHI (M)	*missing value for SHI*	2.9	8.2
PHI	*Private health insurance*	9.8	8.1
no PHI	*1 if they do not have PHI*	9.8	7.8
PHI (M)	*missing value for PHI*	0.5	5.9

Table 13: Demographic variables and percentages of improved health

[28]The choice of this categorisation is analogous to the paper of [Bremer, 2014].

It can be observed in table 13, that although the percentages do not coincide as much as for the values of the descriptive statistics, it does show some similarities. This is because the small differences in the variable's mean and standard deviation from table 11. Moreover, the values in table 13 well reflect those of table 11.

First, as mentioned before that in both samples, respondents in the early half of human life (from age 50 till age 70) which results in more improvements in those age categories. In *sample 2* the percentages of health improvements range between 3.6% and 4.8%, whereas in the other half (from age 70) this percentage is between 1.1% and 2.5%.

As mentioned before 30% and 26% of respondents are located in a small town and in a village, which signifies that more health improvements (6.6% and 6.0% respectively in *sample 2*) can be observed than for those who's main residence is in the city, a large town or is in the suburbs (close to twice as many).

Also as table 11 shows, more than half of both samples have lower education levels, which results in (in table 13) more indications for health improvements, namely 10.8% and 11.4% for *sample 1* and *sample 2* respectively. In the first sample the percentages of improved health for those who have intermediate or high education are only 5.6% and 3.6%.

The last subgroup in table 13 is the insurance status that shows the most dissimilarities. As discussed before, the missing values of these insurance coverages substantially increased in the second sample (due to the limitation in the dataset). This is well reflected in table 13 because for SHI (M) increased from 2.9% (*sample 1*) to 8.2% (*sample 2*), and for PHI (M) this changed from 0.5% to 5.9%. These differences might play a role in the explantation of the differences in the estimation results in section 4.

The next part show these percentages of self-assessed health improvements for the remaining five groups of independent variables in table 14. The correspondence with table 12 can be found as suggested by the previous analysis. Nevertheless, the values in the table below (table 14) match well across samples. There is no significant difference as it could be seen for the missing values of the insurance status.

		Percentage of improved health to total observations	
2) Social Determinants	**Description**	**Sample 1**	**Sample 2**
Family			
alone/single	*single/lives alone*	5.3	4.7
together	*live with spouse/partner*	14.8	17.0
sibling	*1 if ever had any siblings*	17.9	19.7
no sibling	*don't have any siblings*	2.1	2.0
child 0	*not any children*	2.1	2.0
child 1/2	*1 or 2 children*	11.0	12.4
child 3+	*3 or more children*	7.0	7.4
gChild 0	*0 grandchildren*	6.9	7.5
gcChild 1/2	*1 or 2 grandchildren*	5.7	6.1
gChild 3+	*3 or more grandchildren*	7.6	8.1
Social Interactions			
activity	*any activities outside hhd*	9.6	10.1
no activity	*not activities at all*	10.4	11.5
activity (D)	*don't know answer*	0.1	0.1
help given	*gave any help in hhd*	1.2	1.6
no help given	*not any help given*	14.9	16.6
help given (M)	*missing value*	3.9	3.5
help received	received any help in hhd	0.9	1.1
no help received	*no help received in hhd*	7.3	7.4
help received (M)	*missing value*	11.9	13.2
3) Psychological health determinant			
Make ends meet †			
no fdistress	*no financial distress*	12.8	13.6
fdistress	*resp. have financial distress*	7.3	8.1
4) Lifestyle measures			
Health related behaviours †			
no phactive	*no physical activity at all*	2.3	2.5
phactive	*some physical activity*	17.8	19.2
underweight	*BMI <18.5*	0.2	0.3
normal	*18.5 ≤ BMI <25*	7.5	7.9
overweight	*25 ≤ BMI <30*	8.4	9.1
obese	*BMI >30*	3.8	4.1
BMI (D)	*don't know value for BMI*	0.2	0.4
Health evaluations †			
chronic <2	*0 or 1 chronic disease*	10.8	11.6
chronic ≥ 2	*2 or more chronic diseases*	9.3	10.1
no depression	*0-3 EURO-D score*	14.0	15.1
depression	*4 or more EURO-D score*	5.9	6.3
Recent health event †			
hospital	*been in hospital last 12 months*	3.1	4.1
no hospital	*no hospital visit last 12 months*	17.0	17.7
5) Western European Countries †			
country (AT)	*Austria*	2.0	1.5
country (BE)	*Belgium*	4.8	3.9
country (DE)	*Germany*	2.8	2.4
country (ES)	*Spain*	2.3	3.0
country (FR)	*France*	2.6	3.2
country (IT)	*Italy*	2.9	4.3
country (NL)	*Netherlands*	2.8	3.5
6) Other RHS variables			
SAH (M)	*missing values of SAH*	0.0	0.0
OOP	*log of OOP health expenditures*	13.5	13.7

Table 14: Other independent variables and the percentages of improved health

B More Robustness and Sensitivity Analysis

B.1 OOP medical expense as a dummy

The first analysis of this section involves replacing the OOP health expenditures into its participation variant defined in Eq.(5), which essentially tells whether an individual has OOP payment or not. This variable is defined in section 3.1.2 as $\tilde{y}_{2i}$. Note that this dummy variable is equivalent for *model 1* and *model 2* because whenever one has > 0 values for the predictions of $log(y_{2i})$ the other shares the same, thus both results in $\tilde{y}_{2i} = 1$. Table 15 shows the results of this estimations for both samples.

	Coef.	S.E.	M.E.	S.E.	Odds	S.E.
Sample 1	0.017	0.032	0.007	0.009	1.030	0.057
Sample 2	0.060*	0.034	0.018*	0.010	1.110*	0.066

Table 15: Probit estimates, Marginal effects and odds ratios of Eq.(1) for both samples

It can be observed that the results are conflicting with those in the original setting (outcomes from table 8). For *sample 1* they are insignificant (and positive) as with the continuous version and for the second sample, the coefficient shows that individuals who have OOP medical expenditure are more likely to have an improved self-assessed health status. Particularly, its marginal effect is 1.8% and the odds of being in a better state versus not being in a better state 1.11; however these outcomes are only significant on a 10% significance level, which suggest that the estimated effects in the original setting could be more valid but not strong.

B.2 Application of smaller models

This section considers reducing the full baseline models by removing help given or received explanatory variables because they have a considerable amount of missing values in both samples. In *sample 1* there are 19% and 62% of missing values for the help given and received variables respectively, and in *sample 2* the percentages are 18% and 64%. Since respondents have not given and received many helps in the household (0.6% and 0.4% for both samples), all covariants concerning this are removed. That is from table 14 for instance, these are omitted: 'help given', 'no help given', 'help given (M)', 'help received', 'no help received' and 'help received (M)'; therefore, six variables are removed from the original model. Note *model 1* still refers to model with the 2-part estimates on the *first level estimation* and *model 2* to the model with the tobit estimates on that level of estimation. Table 16 shows the estimation results in this reduced set up.

Interestingly, the results in this set up are very similar in comparison to the ones in table 8. Particularly, the standard errors are identical except for the one for the odds ratio in *sample 2* for *model 1*; but that dissimilarity is ignorable as in the original setting it is 0.076 and now it shows 0.077. The main difference between those tables is that the coefficients are not significant on a 1% level anymore but only on the 5% level. This analysis suggests that direction of the effect of OOP payments on the improved health status is clear (negative) but the magnitude is not large.

	2-part model estimates						Heckman	
model 1	Coef.	S.E.	M.E.	S.E.	Odds	S.E.	Coef.	S.E.
Sample 1	0.007	0.095	0.002	0.027	1.01	0.169	0.025	0.028
Sample 2	-0.146**	0.057	-0.043**	0.017	0.779**	0.077	-0.040	0.025

	Tobit estimates						Heckman	
model 2	Coef.	S.E.	M.E.	S.E.	Odds	S.E.	Coef.	S.E.
Sample 1	0.001	0.013	0.000	0.004	1.000	0.022	0.004	0.004
Sample 2	-0.029**	0.011	-0.009**	0.003	0.952**	0.019	-0.007*	0.004

Table 16: Estimation results in the reduced model

B.3 Ignoring Imputed values

As pointed out in the data section, imputed values are used in this investigation to reduce the loss of valuable information and to obtain more efficient estimates [Christelis, 2011]. Another robustness check is done by leaving these values out and observe the change in the nature of the OOP health expenditures' effects; and to do this, imputed observations should be taken out of the samples. Due to the imputation flags that SHARE has created this process can be done straightforward: in a sequential manner, observations are removed if they meet the following selection criteria:

i) imputed value for the dependent variable ($\Delta^{+}SAH$); or if it is an
ii) imputed value for *key explanatory variables*.

Note, the following independent variable contains imputed values: *a*) net income of individual *b*) depression (measured by Euro-d scale) *c*) location of main residence *d*) education level (ISCED code) *e*) number of children *f*) number of grandchildren *g*) and financial distress (if they make ends meet or not).

Therefore, a total of 7,887 and 6,987 observations are imputed for *sample 1* and *sample 2* respectively. This means that the number of observations in this analysis reduces to $11,679 - 7,887 =$ **3,792** for the first sample and $10,574 - 6,987 =$ **3,587** for the second sample.

Table 17 shows the results of the estimations of the *main equation* for this robustness analysis. *Sample 1* has somewhat similar outcomes as the ones presented in the **Results** section. It indicates a much stronger effect and also in the positive direction with marginal effects of 29.8% and 1.1% (although it significantly changed from 0.2% and 0.0% in the original setting) for *sample 1* and *sample 2* respectively. This signifies that *sample 1* does tend to show a positive effect. But the difference of the results of the two models should be noted; again, *model 2* indicates a much less magnitude with smaller (robust) standard errors, which means this model is indeed more robust than *model 1*.

In *sample 2* the two models show conflicting results as in *model 1* the direction of the effect changed to a positive one, while in *model 2* it stayed negative as before. However, since the standard errors in the second model are smaller than that of the first one plus it coincides well with that of the original setting (compare 0.007 and 0.003 from table 8), it can be concluded that *sample 2* does show a negative effect of the OOP medical expense on the indication of improvements in health status.

Model 1

	Coefficient	*S.E.*	*Marginal effect*	*S.E.*
Sample 1	1.09**	0.556	0.298**	0.151
Sample 2	2.65**	1.15	0.775**	0.336

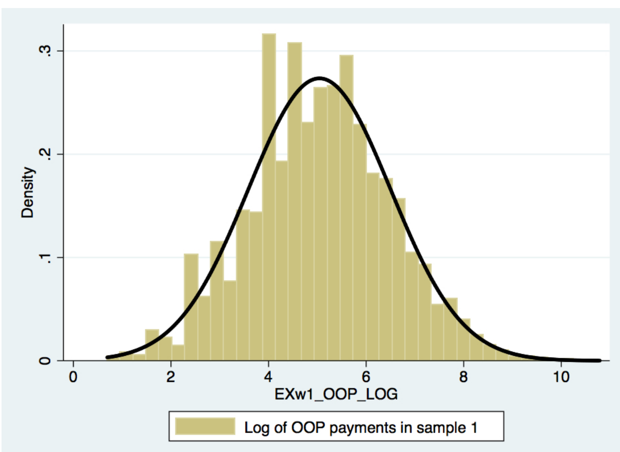
Model 2

	Coefficient	*S.E.*	*Marginal effect*	*S.E.*
Sample 1	0.042**	0.021	0.011**	0.006
Sample 2	-0.060**	0.024	-0.017**	0.007

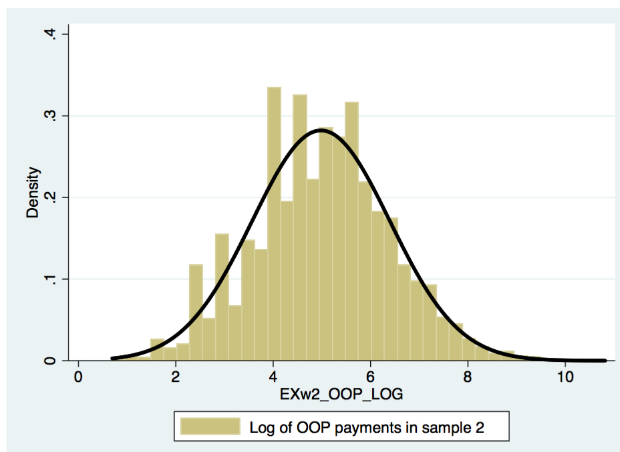
Table 17: Estimation results when imputed values are left out

C Log of Out-Of-Pocket health expenses

The following figures show the histograms of the log of OOP health expenditures, with the normal density function. Figure (1a) is the histogram for the first sample, while figure (1b) is that of the second sample.



(a) Log of OOP payments in sample 1



(b) Log of OOP payments in sample 2

Figure 1: Histogram of the log of OOP payments

It can be well seen that they follow a normal distribution, particularly, the Jarque-Bera test yields a p-value of 0.484 and 0.140 for the first and second histogram respectively. That is, normality of ϵ_i in Eq.(2) cannot be rejected. The kurtosis and skewness of figure (1a) are 2.958 and 0.027, and those of figure (1b) are 0.056 and 2.955 respectively, which resembles the normal distribution's skewness and the kurtosis of 0 and 3. Nonetheless, it is not standard normal.

D Results of the 3-part baseline model (*model 1*)

In this section, tables 18, 19 and 20 show the regression results of the 3-part model (that is the results of *model 1*). Note, the first part (table 18) is the estimation of Eq.(5), the second part (table 19) shows the outcomes of the regression of Eq.(2) and the final part (table 20) estimates Eq.(1).

LHS: $\tilde{y}_{2i}$ in Eq.(5)	Sample 1			Sample 2		
	Coef.	**S.E.**	**P-val**	**Coef.**	**S.E.**	**P-val**
age (55 till 60)	-0.059	0.044		-0.030	0.048	
age (60 till 65)	-0.011	0.049		-0.046	0.054	
age (65 till 70)	-0.188	0.053	***	-0.178	0.058	***
age (70 till 75)	-0.228	0.057	***	-0.182	0.061	***
age (75 till 80)	-0.253	0.061	***	-0.191	0.067	***
age (80plus)	-0.096	0.068		-0.195	0.075	***
male	-0.286	0.031	***	-0.188	0.030	***
education (low)	-0.269	0.044	***	-0.190	0.045	***
education (intermediate)	-0.142	0.044	***	-0.077	0.045	*
education (other)	0.196	0.198		-0.070	0.202	
retired	0.113	0.037	***	0.135	0.038	***
Respondent has PHI	-0.020	0.038		-0.015	0.044	
PHI (M)	-0.182	0.105	*	0.000	0.060	
Respondent has SHI	0.073	0.060		0.236	0.068	***
SHI (D)	0.222	0.134	*	0.305	0.182	*
SHI (M)	1.107	0.068	***	0.209	0.095	**
income (25th percentile)	-0.153	0.046	***	-0.216	0.048	***
income (till median)	-0.119	0.043	***	-0.041	0.045	
income (75th percentile)	-0.071	0.040	*	0.013	0.042	
no financial distress	0.056	0.031	*	0.038	0.034	
been in hospital	0.287	0.040	***	0.259	0.041	***
BMI (normal)	0.203	0.129		-0.089	0.151	
BMI (overweight)	0.229	0.130	*	-0.063	0.151	
BMI (obese)	0.180	0.131		-0.101	0.153	
BMI (D)	-0.109	0.170		-0.277	0.189	
chronic $<$ 2	-0.348	0.028	***	-0.343	0.030	***
no depression	-0.216	0.032	***	-0.129	0.034	***
no physical activity	-0.009	0.046		-0.112	0.049	**
country (AT)	0.899	0.064	***	1.359	0.074	***
country (DE)	1.229	0.062	***	1.462	0.061	***
country (ES)	0.128	0.062	**	0.051	0.060	
country (IT)	0.000	(omitted)		1.185	0.070	***
country (FR)	-0.137	0.053	***	0.153	0.053	***
country (BE)	1.657	0.055	***	1.923	0.061	***
constant	0.188	0.150		-0.060	0.177	

Table 18: Regression results of the first part of the baseline model

In all three tables, the results are shown for both samples, which include the coefficients (Coef.), the robust standard errors (S.E.) and the p-values (P-val). The latter indicates if estimates are statistically significant on a 1% (***), or a 5% (**) or a 10% significance level (*). Additionally, *LHS* stands for the dependent variable in the regressions and *(omitted)* signifies if a variable is omitted due to multicollinearity or not.

The first part is estimated by a probit regression, then given its predicted values the second part can be estimated by an OLS (both with White standard errors). Note, that the regressors in the first two parts are equivalent, while in the third, other health determinants are added. Afterwards, the final part is estimated by another probit regression, given the linear predictions of the second part - that is given $\widehat{log(y_{2i})}$.

LHS: $log(y_{2i})$ in Eq.(2)	Sample 1			Sample 2		
	Coef.	**S.E.**	**P-val**	**Coef.**	**S.E.**	**P-val**
age (55 till 60)	0.052	0.056		0.182	0.062	***
age (60 till 65)	0.135	0.063	**	0.177	0.071	**
age (65 till 70)	0.137	0.069	**	0.223	0.075	***
age (70 till 75)	0.190	0.074	***	0.214	0.079	***
age (75 till 80)	0.266	0.081	***	0.367	0.087	***
age (80plus)	0.604	0.089	***	0.471	0.097	***
male	-0.147	0.037	***	-0.193	0.038	***
education (low)	-0.149	0.051	***	-0.102	0.055	*
education (intermediate)	-0.032	0.048		-0.062	0.050	
education (other)	-0.276	0.280		0.036	0.331	
retired	0.097	0.046	**	0.063	0.048	
Respondent has PHI	-0.083	0.045	*	-0.036	0.055	
PHI (M)	-0.206	0.126		0.131	0.069	*
Respondent has SHI	-0.700	0.118	***	-0.805	0.137	***
SHI (D)	-2.259	0.816	***	0.000	(omitted)	
SHI (M)	-0.029	0.463		-0.853	0.160	***
income (25th percentile)	-0.016	0.057		-0.152	0.061	**
income (till median)	-0.072	0.054		-0.129	0.056	**
income (75th percentile)	-0.086	0.049	*	-0.016	0.052	
no financial distress	0.024	0.039		-0.017	0.042	
been in hospital	0.833	0.044	***	0.805	0.046	***
BMI (normal)	-0.145	0.171		-0.348	0.213	
BMI (overweight)	-0.117	0.171		-0.299	0.213	
BMI (obese)	-0.038	0.173		-0.284	0.214	
BMI (D)	-0.055	0.268		-0.782	0.301	***
chronic $<$ 2	-0.562	0.035	***	-0.567	0.037	***
no depression	-0.254	0.040	***	-0.278	0.044	***
no physical activity	0.395	0.058	***	0.316	0.068	***
country (AT)	-0.010	0.452		0.520	0.322	
country (DE)	-0.123	0.452		0.334	0.320	
country (ES)	-0.337	0.658		0.000	(omitted)	
country (IT)	0.000	(omitted)		0.904	0.325	***
country (FR)	-1.046	0.718		0.205	0.478	
country (BE)	0.590	0.450		0.998	0.323	***
constant	5.876	0.491	***	5.758	0.404	***

Table 19: Regression result of the second part of the baseline model

Although tables 11 and 12 show similarities in the descriptive statistics, tables 13 and 14 present differences in the percentages of self-assessed health improvements as discussed in appendix A. Therefore, these differences can be observed in the results in table 20 as well. In the second sample, more variants are statistically significant than in the first, however, the *sample 1* rejects exogeneity of the financial variable thus parameters are inconsistent.

LHS: $\Delta^{+}SAH$ in Eq.(1)	Sample 1			Sample 2		
	Coef.	**S.E.**	**P-val**	**Coef.**	**S.E.**	**P-val**
Log of OOP payment	0.008	0.095		0.209	0.069	***
age (55 till 60)	0.033	0.044		-0.011	0.059	
age (60 till 65)	0.011	0.051		0.062	0.065	
age (65 till 70)	-0.023	0.058		-0.038	0.072	
age (70 till 75)	-0.131	0.064	**	-0.118	0.077	
age (75 till 80)	-0.013	0.071		-0.090	0.086	
age (80plus)	-0.100	0.092		-0.092	0.096	
male	0.001	0.034		-0.058	0.037	
education (low)	0.045	0.046		0.089	0.052	*
education (intermediate)	-0.010	0.042		-0.018	0.050	
education (other)	-0.153	0.205		0.025	0.212	
retired	-0.041	0.039		0.021	0.043	
location (suburbs)	-0.010	0.053		0.021	0.067	
location (Large Town)	0.041	0.054		0.114	0.067	*
location (Small Town)	0.106	0.049	**	0.149	0.061	**
location (village)	0.009	0.051		0.085	0.063	
Respondent has PHI	0.018	0.039		-0.104	0.047	**
PHI (M)	0.099	0.097		-0.985	0.093	***
Responden has SHI	0.012	0.063		0.084	0.064	
SHI (D)	0.193	0.262		0.129	0.193	
SHI (M)	-0.022	0.073		-1.512	0.306	***
income (25th percentile)	-0.019	0.046		0.000	0.058	
income (meidan)	-0.009	0.044		-0.009	0.052	
income (75th percentile)	0.035	0.040		-0.030	0.048	
give Help	0.018	0.058		-0.109	0.070	
give Help (M)	-0.019	0.066		-0.154	0.082	*
receive Help	0.054	0.074		0.023	0.097	
receive Help (M)	-0.068	0.033	**	0.138	0.041	***
marital status (together)	-0.117	0.057	**	-0.197	0.073	***
sibling	-0.024	0.044		0.012	0.056	
child (1 or 2)	0.071	0.047		0.024	0.059	
child (3plus)	0.095	0.051	*	0.010	0.064	
grand child (1 or 2)	-0.047	0.036		0.053	0.045	
grand child (3plus)	-0.074	0.040	*	0.047	0.047	
activity last 1 month	-0.041	0.029		0.077	0.035	**
activity (D)	0.077	0.181		-0.642	0.349	*
no financial distress	0.022	0.031		-0.035	0.039	
been in hospital	0.046	0.088		-0.307	0.075	***
BMI (normal)	0.090	0.135		0.169	0.154	
BMI (overweight)	0.078	0.135		0.128	0.154	
BMI (obese)	0.054	0.137		0.143	0.156	
BMI (D)	0.007	0.180		0.209	0.215	
chronic < 2	-0.073	0.063		0.319	0.055	***
no depression	-0.073	0.040	*	0.245	0.046	***
no physical activity	0.024	0.061		-0.031	0.063	
country (AT)	0.050	0.069		0.078	0.070	
country (DE)	0.069	0.071		0.121	0.067	*
country (ES)	-0.035	0.091		0.123	0.084	
country (IT)	0.000	(omitted)		1.428	0.295	***
country (FR)	-0.134	0.147		0.031	0.065	
country (BE)	0.000	(omitted)		0.000	(omitted)	
SAH (M)	0.000	(omitted)		-0.752	0.465	
constant	-0.778	0.600		-2.391	0.471	***

Table 20: Regression results of the third part of baseline model

The results of this last regressions (in table 20) has log pseudo likelihood of -5804.8 and -5451.2 for the first and second sample respectively. It can be seen in the table that the obtained values suggest reasonable results.

Considering *sample 2* (because those results show significance in the **Result** section) it suggests that respondents who have their main residence in a town (especially in a small town) are more likely be in an improved health status.[29] This might be rational because the less amount of pollution in the air, or the tranquil environment people tend to feel (thus perceive) an improved health status. Also, in case individuals engaged in any (social) activities outside of the household in the past one month, then the probability of having an improvement in health status by 2010/11 (in comparison to 2006/07) is larger than the probability of staying in the same or worse self-perceived health.

Furthermore, if respondents have been in hospital in the past 12 months, which could be the sign for health problems, then they are more likely to be in the same or even worse health status. They would probably be in the worse self-perceived health, however, due to the binary choice model it cannot be concluded. To investigate such cases, multinomial or conditional logit/probit should be applied for the analysis, which is not carried out in this research but it is an ideal further improvement. Conversely, if respondents have less than two chronic diseases or they have no depression according to Euro-D scale (of 3 or less than 3 scores) in 2006/07 then in 2010/11 they are more likely to have an improvement in their self-assessed health status. This is reasonable because people with many chronic diseases and depression in these older years (50 or above) they would less likely to engage in an improved health status as health generally declines with age.

Last but not least, according to the results presented in table 20, individuals in Italy are more likely to have improvements in health status. However, this statement cannot be supported by any rational motivation. There are also results (missing values of the insurance status) that are significant (even on 1% significance level) but they cannot be well interpreted. Note, this significance in SHI and PHI variants in *sample 2* could be explained by the considerably more number of missing values than in those of *sample 1*.

[29] Note the results corresponds to those of the probit regression, which means that the magnitude of the coefficients cannot be interpreted; only their signs.

References

Balabanova, D. and McKee, M. (2002). Understanding informal payments for health care: the example of Bulgaria. *Health Policy*, 62(3):243–273.

Bellanger, M. and Mossé, P. (2000). Contracting within a centralised health care system: the ongoing french experience. *First meeting of the European Health Care System discussion Group (EHCSDG)*, pages 14–15.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., and Zuber, S. (2013a). Data resource profile: The survey of health, ageing and retirement in europe (share). *International Journal of Epidemiology DOI: 10.1093/ije/dyt088*.

Börsch-Supan, A., Brandt, M., Litwin, H., and Weber, G. (*editors*), (2013b). Active ageing and solidarity between generations in europe: First results from share after the economic crisis. *Berlin: De Gruyter*.

Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist, J., and Weber., G. (2008). First results from the survey of health, ageing and retirement in europe (2004-2007). starting the longitudinal dimension. *Mannheim: Mannheim Research Institute for the Economics of Aging (MEA)*.

Börsch-Supan, A., Brugiavini, A., Jürges, H., Mackenbach, J., Siegrist, J., and Weber, G. (2005). Health, ageing and retirement in europe – first results from the survey of health, ageing and retirement in europe. *Mannheim: Mannheim Research Institute for the Economics of Aging (MEA)*.

Börsch-Supan, A. and Jürges, H. (*editors*), (2005). The survey of health, ageing and retirement in europe – methodology. *Mannheim: Mannheim Research Institute for the Economics of Aging (MEA)*.

Bremer, P. (2014). Forgone care and financial burden due to out-of-pocket payments within the german health care system. *Health Economics Review*, 4(1):1–9.

Calltorp, J. and Abel-Smith, B. (1994). Report on the greek health services. *Pharmetrica*.

Chawla, M., Berman, P., and Kawiorska, D. (1998). Financing health services in poland: New evidence on private expenditures. *Health Economics*, 7(4):337–346.

Christelis, D. (2011). Imputation of missing data in waves 1 and 2 of share. *Available at SSRN 1788248*.

Crystal, S., Johnson, R. W., Harman, J., Sambamoorthi, U., and Kumar, R. (2000). Out-of-pocket health care costs among older americans. *Journal of Gemntology: SOCIAL SCIENCES*, 55(1):S5I–S62.

Denton, M., Prus, S., and Walters, V. (2004). Gender differences in health: a canadian study of the psychosocial,structural and behavioural determinants of health. *Social Science and Medicine*, 58(12):2585–2600.

Dixon, A., Figueras, J., and Kutzin, J., editors (2002). *Funding health care: options for Europe*. Buckingham: Open University Press.

Dormont, B., Grignon, M., and Huber, H. (2006). Health expenditure growth: reassessing the threat of ageing. *Health Economics*, 15(9):947–963.

Greene, W. H. (2002). *Econometric Analysis.* Upper Saddle River, New Jersey, Upper Saddle River, New Jersey, 07458, fifth edition.

Heij, C., de Boer, P., Franses, P. H., Kloek, T., and van Dijk, H. K. (2004). *Econometric Methods with Applications in Business and Economics.* Oxford University Press Inc., New York.

Heijink, R., Xu, K., Saksena, P., and Evans, D. (2010). Validity and comparability of out-ofpocket health expenditure from household surveys: A review of the literature and current survey instruments. *World Health Organization.*

Johnston, D. W., Propper, C., and Shields, M. A. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, 28(3):540–552.

Jürges, H. (2006). Does ill health affect savings intentions? *MEA Discussion Paper No. 139-07.*

Kim, D. and Whasington, S. (2006). The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis and Prevention*, 38(6):1094–1100.

Leung, S. and Yu, S. (1996). On the choice between sample selection and two-part models. *Journal of Econometrics*, 72(1):197–229.

Lyseight, L. (2010). *1001 Life Changing Quotes 4 TEENS: Chart Your Success Path with Wisdom words.* Xlibris, Corp.

Malter, F. and Börsch-Supan, A. (*editors*) (2013). Share wave 4: Innovations and methodology. *Munich: MEA, Max Planck Institute for Social Law and Social Policy.*

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, 365(9464):1099–1104.

Marshall, S., McGarry, K., Skinner, J. S. and Wise, D. A (*editor*), (2011). The risk of out-of-pocket health care expenditure at end of life. *Explorations in the Economics of Aging. University of Chicago Press*, pages 101–128.

Nordhaus, W. D. (2005). The health of nations: The contribution of improved health to living standards. *American Journal of Economics and Sociology*, 64(8818):367–392.

Rubin, R. M. and Koelln, K. (1993). Determinants of household out-of-pocket health expenditures. *Social science quarterly*, 74(4):721–735.

Sambamoorthi, U., Shea, D., and Crystal, S. (2003). Total and out-of-pocket expenditures for prescription drugs among older persons. *The Gerontologist*, 43(3):345–359.

Schieber, G. J., Poullier, J.-P., and Greenwald, L. M. (1992). U.s. health expenditure performance: An international comparison and data update. *Health Care Financing Review*, 13(4):1.

Seshamania, M. and Gray, A. (2004). Ageing and health-care expenditure: the red herring argument revisited. *Health Economics*, 13(4):303–314.

Stewart, S. T. (2004). Do out-of-pocket health expenditures rise with age among older americans? *The Gerontologist*, 44(1):48–57.

Szende, A. and Culyer, A. J. (2006). The inequity of informal payments for health care: The case of hungary. *Health Policy*, 75(3):262–271.

Unesco (1997). *International Standard Classification of Education-ISCED 1997: November 1997*. Unesco.

WHO (2003). *Atlas of health in Europe*. World Health Organization. Regional Office for Europe.

Windmeijer, F. and Silva, J. S. (1997). Endogeneity in count data models: an application to demand for health care. *Journal of applied econometrics*, 12(3):281–294.

Winkelman, R. (2003). *Econometric Analysis of Count Data*. Springer Science and Business Media.

Wissen, L. J. and Golob, T. F. (1990). Simultaneous equation systems involving binary choice variables. *Geographical Alnalysis*, 22(3):224–243.

You, X. and Kobayashi, Y. (2011). Determinants of out-of-pocket health expenditure in china. *Applied health economics and health policy*, 9(1):39–49.

Online sources:

Giles, David E. 'Robust Standard Errors For Nonlinear Models'. *Econometrics Beat: Dave Giles' Blog 2013*. Retrieved from `http://davegiles.blogspot.hu/2013/05/robust-standard-errors-for-nonlinear.html`
Last accessed: 5 July, 2015.

SHARE Release Guide 2.6.0 of waves 1&2:
`http://www.share-project.org/fileadmin/pdf_documentation/SHARE_guide_release_2-6-0.pdf`
Last accessed: 29 June, 2015.

Constitution of WHO (1946) on the importance of health:
`http://www.who.int/trade/glossary/story046/en/`
Last accessed: 29 June, 2015.

More information on Anne Wilson Schaef:
`http://www.livinginprocess.com/anne-wilson-schaef.php`
Last accessed: 29 June, 2015.

The picture of Erasmus on the title page:
`http://www.thereformation.info/Images/Erasmus.jpg`
Last accessed: 5 July, 2015.

The (small) Erasmus signature:
`http://repub.eur.nl/eur_signature.png`
Last accessed: 5 July, 2015.

The (large) Erasmus signature in the background:
`http://www.porteconomics.nl/EUR%20handtekening.gif`
Last accessed: 5 July, 2015.