The relationship between unemployment and health

Katalin Gaspar

Central European University
Department of Economics

In partial fulfillment of the requirements for
the degree of Masters of Arts

Supervisor: Prof. Gábor Kőrösi
Submitted: 2009, Budapest, Hungary

Abstract

Using a Hungarian longitudinal health survey I analyze the causal relationship between unemployment and health, and I examine whether any significant relationship exist after simultaneity is treated. After establishing the presence of simultaneity, I use three different methods to circumvent the problem. The main conclusion throughout the four models is consistent. Unemployment has a detrimental effect on health, and vice versa poor health will lead to a higher probability of unemployment after endogeneity is treated.

I find that the length of unemployment will influence the change in health-status between the two surveys. The deterioration in health is the greatest for those who were unemployed during both surveys. Less health deterioration can be measured among those who transitioned from unemployment to work during the two surveys, and the least amount of deterioration is noticeable among those who transitioned from work to unemployment. Examining the marginal effects reveals that employment transitions are the most significant in explaining a change in health if the change in health-status is moderate, which means that transitions will not lead to extreme changes in health.

Contents

List of Tables

Introduction

A great deal has been written in recent years on the effects of inactivity on health, and on the effects of unemployment on health. Most relevant papers conclude that the relationship is negative, however it is still unclear what the direction of causality is between the two and whether there is any real link after endogeniety is treated. In this study I will focus on unemployment. In particular, I will evaluate whether unemployment has a significant effect on health after fixing the endogeneity problem, and whether this effect is substantial enough to validate new policy. Finding the true effects of unemployment on health could be critical both in terms of labor policy and in terms of health-policy formation. In addition to the human costs, health deterioration also will burden public health expenditures and further diminish labor supply. Hence, in order to facilitate employment, policy-makers must develop tools that moderate the detrimental effects of unemployment on health.

This paper is organized as follows: in Chapter 1 I briefly discuss the theoretical framework behind the endogeneity problem between unemployment and health. In Chapter 2, I review relevant literature. In Chapter 3, I discuss the four different models I use to analyze the dynamics of the relationship. I present my data and my findings in Chapter 4 and 5, and I conclude in Chapter 6.

Chapter 1: Theoretical framework

Theoretical grounding for the topic is usually originated from the work of Currie and Madrian (1999). According to their model consumers maximize a utility function, which in addition to consumption and leisure also includes health. Individual must invest time and health goods into staying healthy: when ill the consumer loses time working and loses wages that could be spent on consumption and leisure. On the other hand, the consumer must spend time working to be able to purchase health goods.

Poor health will influence the marginal rate of substitution between leisure and labor by increasing the “marginal disutility from labor”. That is the poorer one’s health, the less s/he is willing to supply labor. Hence, the underlying endogeneity in the model comes from the fact that labor supply is a function of health while health is a function of the amount of labor supplied and the amount of wages received. (Gordo, 2006).

Chapter 2: Literature Review

The relationship between unemployment and health has been a widely studied area of economics. Initial research began with time-series analysis of aggregate data, evaluating the correlation between national (or regional) output, unemployment-rates and mortality-rates. Brenner et al. (1983) examined Western countries, including England and Wales from 1936 to 1976, and the United States from 1909 to 1976. He concluded that there is a positive correlation between overall unemployment and mortality-rates. On the other hand, Adams (1981) established a negative correlation between unemployment and mortality based on a Canadian time-series analysis between 1950 and 1977.[1]

Instead of analyzing aggregate time-series variables, more recent studies have turned to longitudinal data. Moser et al. (1984) used a longitudinal survey to evaluate the mortality-rate of men between the ages of 15 to 64, who were seeking employment one week prior to the survey. He found that jobseeker men had mortality ratios that were higher than the average standardized rate. Some of the gap disappeared after controlling for socio-economic status (SES), but close to 20-30% of the difference remained. The study did prove a strong correlation between unemployment and health, but it did not specifically evaluate what factors of unemployment

[1] The author mentions a few possible explanations for his nonconforming results. He suggests " 1) faulty assumptions concerning lag times between increases in unemployment and increases in mortality rates, 2) dilution of the association between joblessness and actual financial hardship by unemployment insurance benefits, 3) a decrease in societal [business] activity as a result of unemployment , and therefore lower risks of work-related deaths 4) reduced alcohol and tobacco consumption because lack of employment income, and therefore lower mortality risks, and 5) confounding of the relation between unemployment and adverse health outcomes because of inequality in income. " (p. 531, Jin 1995)

caused this effect. Interestingly, Moser also found higher mortality-rates among women married to the job-seeker men. The findings of this study were robust for the 1971-75 as well as for the 1976-81 periods although he did not explicitly control for endogeneity between unemployment and health.

Another study by Lavis published in 1998 analyzed local unemployment rate, the duration of unemployment and mortality on longitudinal data. He found that men who had one or more spells of unemployment had higher hazards of death than those who were working or were in retirement. He also found a connection between the length of unemployment and death. He observed that men who experienced longer unemployment spells died earlier than those who experienced shorter or no spells of unemployment. However, he found no clear relationship between the number of unemployment spells and mortality.

We can obtain more insight into the mechanics of the relationship between unemployment and mortality by comparing the general health status of individuals with different employment histories. The usual variable of study for such analysis is the self-rated health (SRH) indicator (see Chapter 3 for more detail on this variable). The advantage of using this variable is that it enables the researcher to quantify changes in health and analyze the underlying cause of the change. For example, Ahs (2006) used Swedish longitudinal data to analyze the fluctuation in the gap between the average health status (measured by SRH) of employed and unemployed individuals over two five year periods. After adjusting for SES factors, she found that the general health status of unemployed is lower in times of high unemployment, than in good economic times.

Another convincing piece of evidence on the topic comes from a study on the link between the loss of job-security and health. Ferrie (1995) studied a group of British civil servants going through privatization, and found that even the potential loss of employment has an adverse effect on SRH and that health status is not completely restored when job security is reestablished. She found that men exhibit clear signs of deterioration in SRH, while in the case of women this effect is more restricted.

Researchers using longitudinal data have been able to employ different econometric techniques in order to control for the endogeneity in the model. Cooper et al. (2007) used duration analysis to analyze the effects of unemployment on the duration of good health and found a negative relationship but to different degrees between men and women. The authors found that the health deterioration of unemployed men is considerably larger than it is for women. They argue that this difference comes from the male gender role as the breadwinner of the family, which leads to more stress for unemployed men compared to unemployed women who have an alternative of household labor.

Gordo (2006) used panel data analysis on German longitudinal survey to analyze how the duration of unemployment affects health deterioration. She found that short-term unemployment is only detrimental to men, but not for women. On the other hand, when someone is unemployed for more than 2 years, health deterioration is significant in the case of both men and women. In addition, Gordo also evaluated the effect of re-employment on health and found that becoming employed again has a positive effect on health for both men and women independent of the duration of unemployment.

There is another strand of literature that deals with the deterioration in health among the general public in ex-Socialist states. Most papers in the category attempt to explain the sudden increase in mortality-rates that was apparent in nearly all of the transition economies after the political change took place in 1989. The case of Central and Eastern European countries is particularly interesting, since mortality-rates prior to its dramatic rise in these countries were especially low even in an international comparison. Kopp (2005) states that mortality rates in the 1970s in Hungary were lower than those observed in Great Britain or Austria. However, trends in the East and the West diverged due to the transition. As Western European mortality rates declined, Eastern European rates rose back to their levels in the 1930s and remained there for the next decades.

There is a wide range of explanations used to justify this phenomenon, however I will only focus on studies that evaluate the effects of mass layoffs and economic contraction on the general health status of the population. For example, Demirchyan (2008) and co-authors evaluated the determinant of poor SRH of Armenian women on cross-sectional surveys from 2001 and 2004. They found that SES variables such as moderate to severe material deprivation, level of education, as well as health indicators such as smoking were the major determinants of poor health. These factors mediated any effects women's economic status and access to healthcare may have had. Bobak (1998) analyzed the determinant of SRH in Russia also in a cross-sectional setting with similar results as above. The researchers found that „perceived control over life" was one of the main factors in SRH, along with material deprivation and education. The study also found significant evidence for the prior belief that poor health is somehow correlated with political background; evidence shows that subjects who opposed the economic transition were in poorer health than the rest of the population.

Hungarian research on the broader topic of socio-economic status and health includes a cross-sectional study conducted by Tahin et al. (1999) who examined the relationship between an individual's level of education and health, which found a strong positive correlation between one's number of years of studies and health status after controlling for gender, age, marital status. However, they argue that this correlation is indirect, since financial stability, better work environment and a generally higher standard of living due to higher education are the main factors that lead to better health. Kopp et al. (2004B) evaluated whether the marital status of individuals and the SES of the partner has an effect on one's health. She found that both sexes are affected, however wife's SES has a greater effect on the husband's health status than vice versa.

Chapter 3: Model

The main difficulty in estimating the health effect of unemployment is the underlying endogeneity between one's health status and labor force status. Poor health and unemployment may be jointly determined in the model. An individual may become unemployed due to an illness or become ill as a result of becoming unemployed. In general, one can establish four types of causality:

(1) Unemployment $\rightarrow$ Health

(2) Health $\rightarrow$ Unemployment

(3) Health $\leftrightarrow$ Unemployment

(4) Correlation through an exogenous variable or spurious correlation

In case (1), a spell of unemployment leads to deterioration in health in the future. In case (2), poor health leads to unemployment though for example high absenteeism, low productivity of an employee etc. In case (3), health and unemployment are jointly determined. That is, both have significant effects on the other in the same period, this is a case of simultaneity. And in case (4), no real causal relationship exists between the two variables. Their perceived link is a result of their correlation with another exogenous variable, or a result of a spurious correlation. Since health and unemployment can be explicitly controlled for, case (1) and (2) will not lead to endogeneity in the model. And since in case (4) the causality is only perceived, endogeneity will not be present here either. It is only in case (3) that endogeneity in the form of simultaneity is present.

Section 1 : Regression with lags

In order to establish the presence of simultaneity, I regress both SRH and unemployment on their own lags and on the lags of the other endogenous variable and evaluate their significance. I will estimate the following equations:

$$h_{it} = X_{it}\beta_t + u_{it} + u_{it-1} + h_{it-1} + e_{it}$$

and

$$u_{it} = X_{it}\, \beta_t + h_{it} + h_{it-1} + u_{it-1} + e_{it}$$

where $X_{it}\,\beta_t$ are exogenous individual regressors, h_i is SRH and u_i refers to the state of unemployment. If at a given period "t" h_i and u_i are both significant in the two equations, then either case (3) or case (4) hold. That is, either simultaneity or correlation through an exogenous variable is present. However, if I prove that simultaneity is present I also show that the coefficients in this regression are biased. If h_i and u_i are jointly determined, then they will both be correlated with the error-term, leading to an upward bias in absolute value in the coefficients. After establishing the presence of simultaneity, I provide three possible techniques for eliminating this bias. In Section 2, I use OLS on differences to eliminate the individual effects along with the simultaneity. In Section 3 and in Section 4, I use two variants of the instrumental variable regression to circumvent the endogeneity problem.

Section 2: Regression on differenced variables

In this model I use simple OLS model on differences with binary variables representing transitions in and out of employment. I evaluate how transitions influence the change in SRH after controlling for individual characteristics. To illustrate the method, assume that the health-equations looks like the following:

$$h_{it} = X_{it}\beta_t + \sum_{1}^{4} d_{it}^{\,u} + \nu_i + e_{it}$$

The equation at this state includes time-invariant variable ν_i, which incorporates all unobservable heterogeneity invariant in time including personal characteristics, such as economic background, prior history of unemployment as well as prior history of health. We know that endogeneity and the bias in the OLS coefficients comes from the correlation between individual characteristics ν_i and e_{it}. We also know that this correlation may be present even without the existence of simultaneity, as individual characteristics in the exogenous variables (now incorporated in ν_i) will correlate with the error-term and lead to a bias. We can remove (or at least diminish) the simultaneity (and endogeneity) bias by eliminating this time-invariant term by subtracting one period from the other. However, this comes at a cost. Taking differences will not only remove ν_i, it will also eliminate a great portion of the dynamics in the model.

Section 3.: IV Regression

In Section 3, I employ a 2SLS model in order to fix the endogeneity problem. The appropriate instrument will be strongly correlated with the endogenous variable, but will be uncorrelated with the error term. For example, take the two equations as before:

$$h_{it} = X_{it}\beta_t + u_{it} + u_{it-1} + h_{it-1} + e_{it}$$

and

$$u_{it} = X_{it}\beta_t + h_{it} + h_{it-1} + u_{it-1} + e_{it}$$

The endogeneity comes from the fact that u_{it} and u_{it-1} in the health-equation and h_{it} and h_{it-1} in the unemployment-equation are correlated with e_{it}. Therefore, if I find a set of instruments that are strongly correlated with the endogenous variables, but are not correlated with the error term e_{it} I can circumvent the simultaneity problem. However, finding such an instrument is difficult. In my case, instrumenting health is straightforward. I use BMI and smoking as IVs. On the other hand, finding a valid instrument for unemployment is not as simple. I present my outcomes using education and regions as possible instruments. However, these variables do not meet all the conditions of a valid IV (individually as well as jointly). Results and problems are presented in Section 3.

Section 4: Simultaneous Regression

In this section, I use another variant of the IV method used in Section 3, with the only difference that the two equations are estimated in a simultaneous manner. My instruments will be the same as above. I use BMI and smoking to instrument SRH and education and region to instrument unemployment. As before the general conditions for the validity of an instrument must be met, which in my case is not completely satisfied.

Chapter 4: Data

Hungarostudy Health Panel (HEP) is a Hungarian longitudinal survey specifically designed for measuring the change in the health status of individuals over time. The survey includes a thorough questionnaire regarding individual's mental and physical health status, but it is rather weak on socio-economic variables. The HEP-survey was conducted twice in 2002 and 2005-2006, and 4524 people took part in both waves. In this study, I only use data for active or unemployed individuals present in both waves of the survey between the ages of 25 and 55. Descriptive statistics are presented in the Appendix.

Self-rated health

For measuring health-status I will use the self-rated health (SRH) indicator, which in the HEP-survey is an ordered variable from 1 to 5 (1 being very ill and 5 in perfect health). The reliability of SRH is a hotly debated issue among empirical researchers. By using this subjective measure of health to proxy a latent "true" state of health, one implicitly assumes that the SES-factors that influence health will have a similar effect on the individual's subjective perception of his/her health-status. However, Simon (2005) found that higher educated people tend to report aspects of wellbeing (eg. happiness, their ability to cope with an illness etc.) as part of health status, while those with lower levels of education restricted answers to physical aspects of health. Huisman et al. (2007) evaluated the ability of the SRH in predicting mortality-rates among individuals with different levels of education. The authors concluded that although there seems to be a positive correlation between the level of education and the ability of the indicator to predict mortality, this effect is small. Therefore studies using this indicator will not overestimate true health.

Table 1: Change in SRH according to employment transitions and gender

		TRANS1		TRANS2		TRANS3		TRANS4		
		t-1: work, t: unempl.		t-1: unempl., t: unempl.		t-1: work, t: work		t-1: unempl., t: work		
		Male	female	male	Female	male	Female	male	female	Total
N		64	40	35	25	674	731	46	30	1645
Avg. Change in SRH	>	22%	15%	31%	4%	11%	12%	20%	23%	211
	=	39%	45%	29%	52%	39%	40%	41%	37%	657
	<	39%	40%	40%	44%	49%	48%	39%	40%	777
		100%	100%	100%	100%	100%	100%	100%	100%	

Table 1 illustrates the four transitions tabulated against the change in SRH between the ages of 25 to 55. Change in SRH (ΔSRH) is calculated by subtracting SRH obtained in 2005 from than obtained in 2002. The share of those with a positive change in SRH is the largest among individuals who were working during both surveys, but it is the lowest among those who had a transition between the two surveys. Also, there is a striking gender difference in health satisfaction among the long-term unemployed (TRANS2): while health deterioration (ΔSRH<0) was indicated by 31% of men, the same number was only 4% for women. Consistently, no change (ΔSRH=0) and improvement in health (ΔSRH>0) were indicated by significantly more women than men. Table 14. in the Appendix presents SRH tabulated against average age.

Education

Huisman (2007) proves a significant correlation between an individual's education and the chosen SRH's ability to predict future mortality. As the level of education increases, so does an individual's ability to take into consideration the multi-dimensionality of his/her health status. This means for example that out of two people with the same level of SRH measure the person with a lower level of education will have a higher probability of mortality than someone with a higher level of education.[2] Hence, the inactive and unemployed who are also more likely to have lower levels of education will most likely underestimate their own probability of mortality.

[2] Huisman (2007) found in his data that „subjects with primary education had a higher crude risk of dying if they indicated that their health was 'very good' (12.3% in primary educated, 3.8% in tertiary educated), but they had a lower crude risk of dying if they had indicated that their health was 'bad' (33,8% in primary, 35,3% in tertiary)."

Unemployment Status

Economic activity of respondents was registered in both surveying waves. Unfortunately a common set of definitions for each of the answer choices was not established. This will undoubtedly influence the results, as survey conductors and respondents could have varying interpretation of each of the answer choices. In the case of unemployment this is particularly crucial: the generally accepted definition of an unemployed is someone who has not been in paid employment during the reference period and who is actively looking for employment[3]. Therefore, those individuals, who have been out of work for a longer period of time but have not been able to find work and hence have given up on finding work should be considered inactive or out of the labor force, and not as unemployed. In the case of the HEP-survey, I will assume that the unemployed will include not only those who are not employed and are actively looking for work, but also those who would like to work but have given up on searching for employment. In other words, all those who do not have an alternative reason for staying at home (such as pensioners, students, those on maternity leave, or simply dependent etc.) as unemployed.

Wealth Indicators

Due to changes in income-related questions between the two survey waves, I am restricted to using a subjective wealth indicator to measure wealth and income. Similarly to the SRH, the *subjective wealth indicator* is an ordinal variable between 0 and 10 (10 being the highest level of wealth). Unlike the SRH variable, the “subjective wealth” variable has scarcely been researched, but it can be assumed that individual characteristics such as one’s age, education, and region will influence his/her opinion of own subjective wealth.

[3] According to the official ILO definition „the unemployed comprise all persons above a specified age who during the reference period were: (1) without work, that is, were not in paid employment or self employment during the reference period. (2) currently available for work, that is, were available for paid employment or self-employment during the reference period; and (3) seeking work, that is, had taken specific steps in a specified recent period to seek paid employment or self-employment. (OECD Glossary)

Married-variable

Independent of their official marital status, the married-variable equals 1 for anyone who indicates that they live in the same household with (1) partner or (2) married spouse.

Age

Age is restricted between 25 and 55 years in order to exclude students and pensioners. The risk of mortality naturally increases with one's age, independent of SES and other lifestyle factors. By restricting my analysis to those below 55 I have ruled out those most at risk of premature mortality. In my estimations below, I add age divided by 10.

BMI

The body mass index is a statistical measure that helps healthcare professionals in determining an individual's healthy body weight, based on his/her weight, height and hip-size. According to the scale an individual can be characterized as underweight, normal, overweight or obese.[4][5]

Smoking and gender

Smoking and female variables are entered into the estimation in binary form.

[4] The general formula for the BMI-index is: BMI= weight (kg) / $height^2$ (meters2).

[5] Individuals with BMI under 18.5 are regarded at underweight, between 18.5 and 25 as normal, between 25 and 30 as overweight, and above 30 as obese.

Table 2. List of variables

Dependent vars	Type	Detail
SRH	5 level ordered	(1) poor (2) fair (3) average (4) good (5) excellent
ΔSRH	Ordered	differenced SRH
Unemployed	binary variable	'=1 unemployed
Independent vars		
Married	binary variable	=1 if lives with partner/spouse
AGE	continuous variable	25-55 years of age, divided by 10
EDU1	binary variable	Less than 8 years of education
EDU2	binary variable	8 years of education
EDU3	binary variable	Vocational training (szakmunkásképző)
EDU5	binary variable	Technical secondary school (szakközép érettségi)
EDU6	binary variable	General secondary school (gimnáziumi érettségi)
EDU7	binary variable	College or university diploma
Subj Wealth	0-10 level ordered	
Unemployed	binary variable	=1 if unemployed
City	binary variable	=1 if lives in city
BMI1	binary variable	=1 if BMI below 18.5
BMI2	binary variable	=1 if BMI between 18.5 and 25
BMI3	binary variable	=1 if BMI between 25 and 30
BMI4	binary variable	=1 if BMI above 30
Smoking	binary variable	=1 if smoker
Female	binary variable	=1 if female
Regions	canonical variable	see Appendix for details
TRANS1	binary variable	=1 if t-1: worker & t: unemployed
TRANS2	binary variable	=1 if t-1: unemployed & t: unemployed
TRANS3	binary variable	=1 if t-1: worker & t: worker
TRANS4	binary variable	=1 if t-1: unemployed & t: work
KIDS	binary variable	=1 if small child in the household
ΔSRH1	binary variable	=1 if ΔSRH1<-1
ΔSRH1	binary variable	=1 if ΔSRH1=-1
ΔSRH1	binary variable	=1 if ΔSRH1=0
ΔSRH1	binary variable	=1 if ΔSRH1=1
ΔSRH1	binary variable	=1 if ΔSRH1>1

Chapter 5: Estimation

In Section 1 of this chapter I present a simple OLS model with lags in order to confirm the presence of simultaneity. Next in Sections 2, 3, and 4, I introduce three techniques for circumventing endogeneity. I display both OLS estimation as well as probit and ordered-probit results for all regressions.[6]

In general, I anticipate that the coefficients on two endogenous variables will be negative and significant. However, I expect that in Section 1 these coefficients will be larger in absolute value than in the following three sections. The correlation between SRH and unemployment is negative, therefore the partial effect between unemployment and health should decline after simultaneity has been controlled for. In addition, in Section 2 coefficients will decrease in size due to the loss of dynamics caused by the differencing.

Section 1. : Regression with lags

OLS and probit results with lags are shown in Table 3 below. I use this model to determine whether simultaneity is present in the model. Looking at the health equation we see that current unemployment and past unemployment are both negatively correlated with health and so is past SRH. In the unemployment-equation, current SRH and past unemployment look significant, but again their significance is higher in the probit setting which indicates that the correlation is not linear. When evaluating the marginal effects of the ordered-probit estimation (Table 18 in the Appendix), we see that BMI and smoking will significantly influence health at every SRH-level. The probability of choosing SRH1 through SRH3 will increase by up to 10.5% if the individual has high BMI (BMI3 and BMI4) depending on the SRH chosen. Same holds for smoking: if someone is a smoker s/he is more likely to choose SRH1 through SRH3 than anything higher.

[6] Unemployment is a binary nominal variable, while SRH has an ordered nature. Although OLS is not biased when used on nominal or ordered variables, it loses efficiency. In this case, ordered-probit (or logit) maximum-likelihood estimation is a more efficient estimator. It is assumed that the error terms are normally distributed. However, when it is unknown whether the dependent variable is ordered or not, it is safer to use a least-squares model, since using an ordered-type model on a non-ordered dependent variable will introduce a bias. (Gatignon, p. 147) Therefore I present both OLS as well as maximum likelihood estimates.

Surprisingly, subjective wealth has only a limited significance in affecting the individual's SRH choice, the variable is only significant at 1% at SRH5, and the size of coefficient is only a fraction of those on the BMI's and smoking. In addition to the health variables, unemployment and past SRH also play a major role in determining health status. Understandably, past SRH has a relatively high effect on current health when compared to other marginal effects in the estimation. However, it is lower than expected: past SRH will only increase the current probability of choosing SRH4 and SRH5 by 13% and 5%, respectively. This indicates that there is a substantial amount of variation between the registered SRH during the two survey periods (I will use this fact in Section 2. to treat endogeneity).

All other SES exogenous variables, such as gender, type of habitation, family status, and having children in the home, seem to be independent of health at every SRH-level. This is a surprising result, since these factors (especially gender and family status) were found positive in other studies on the topic. [7]

Past unemployment seems to be the strongest predictor of current unemployment when analyzing the marginal effects in the unemployment-equation (Table 19 of the Appendix). Lagged unemployment adds a 26.5% probability to being unemployed today. In addition, education also has a considerable effect. Having a diploma from technical secondary school or higher will decrease the probability of unemployment by more than 10%. The coefficient on SRH is also strongly significant, which proves that on average poor health leads to unemployment. However, the size of the coefficient is small when compared to education variables. One unit increase in SRH will lead to around 3% decrease in the probability of unemployment, and while only significant at 10%, lagged SRH will also cause around 2% increase in the probability of being out of work.

[7] Kopp(2004B) for example found great differences in the mortality-rates of men and women in Hungary between the ages of 45 and 64. She also found that female level of education will only slightly influence female mortality, while it will significantly influence male mortality. Although I use SRH and not mortality as the dependent variable, I find that education does not significantly influence health.

Based on these results I can conclude that simultaneity is indeed present in the model, since both current period endogenous variables are significant in the linear as well as in the non-linear estimation. Also, it seems that the simultaneity is somewhat skewed towards the health effect of unemployment. Past and current unemployment has a *strong* effect on health, while health has only a *limited* (but significant) effect on unemployment.

Out of the possible directions in causality presented in Chapter 2, either Case (3) or Case (4) holds. That is, the two endogenous variables are either simultaneously determined or they are both influenced by another exogenous variable not presently controlled for in the equations. If I assume that my specification is correct and all observable individual effects are controlled for, I only need to consider time-varying factors as the possible exogenous source of this correlation. Economic fluctuations could play such role: for example if 2004-2005 were particularly bad years in economic output, this would influence unemployment and health *independently,* without one variable determining the other.[8] However, since 2004-2005 were not exceptionally weak in terms of the economy and no other time-varying factor could have the same direction of effect on both variables I will rule out Case (4), and conclude that Case (3) is the valid direction of causality, i.e. the variables are determined endogenously in a simultaneous manner. On the other hand, simultaneity in the model means that the coefficients presented in Table 3 are biased. Therefore, although the strong significance on the lagged unemployment variable in the health-equation indicates that there is a delayed relationship leading from unemployment to health (in addition to the simultaneity) this cannot be proven for certain.

Nonetheless, based on this initial (albeit biased) result we can conclude that BMI, smoking, and to limited extent subjective wealth and age will have some significant effect on health. And

[8] As discussed in the Literature Review in Chapter 2, not only becoming unemployment but any work related stress especially the loss of job security will have a detrimental effect on health.

conversely, age, education and to limited extent regional differences will be important in determining unemployment. [9]

[9] Initially I used age variables in binary form by decades in all regressions. However this did not sufficiently control for the effects of age. Variables that were assumed to be also correlated with age (such as married and kids) had irrational signs.

Table 3: Health Equation - Estimation with Lags

	OLS		Ordered probit	
SRH	Coef.	S. E.	Coef.	S. E.
AGE	-0.037	0.232	-0.092	0.391
AGE2	-0.022	0.029	-0.034	0.048
BMI4	-0.111***	0.042	-0.195***	0.072
BMI3	-0.203***	0.052	-0.345***	0.087
Smoking	-0.157***	0.039	-0.275***	0.066
Subj. Wealth	0.033**	0.013	0.052**	0.022
Female	-0.050	0.038	-0.095	0.064
Married	-0.018	0.048	-0.034	0.082
City	-0.018	0.037	-0.025	0.063
Kids	-0.011	0.044	-0.020	0.075
Unemployed	-0.164**	0.065	-0.278***	0.108
Lagged Unemployed	-0.186***	0.068	-0.308***	0.113
Lagged SRH	0.287***	0.030	0.477***	0.052
Constant	3.329***	0.449		
N	1295		N	1295
F 13, 1281	23.19		Wald chi213	258.29
Prob > F	0.0000		Prob > chi2	0.0000
R-squared	0.2063		Pseudo R2	0.1033
Root MSE	.65922			

Unemployment Equation - Estimation with Lags

	OLS		Ordered probit	
Unempl.	Coef.	S. E.	Coef.	S. E.
AGE	0.014	0.115	-0.026	0.647
AGE2	-0.004	0.014	-0.014	0.080
EDU2	-0.188	0.174	-0.584	0.543
EDU3	-0.267	0.172	-0.887	0.540
EDU4	-0.308*	0.173	-1.170**	0.554
EDU5	-0.316*	0.173	-1.240**	0.559
EDU6	-0.325*	0.173	-1.520***	0.581
Subj. Wealth	-0.011**	0.005	-0.060*	0.033
Female	-0.007	0.017	-0.035	0.110
Married	-0.023	0.022	-0.146	0.129
REGION2	0.020	0.028	0.170	0.259
REGION3	0.043	0.031	0.315	0.258
REGION4	0.036	0.028	0.297	0.251
REGION5	-0.016	0.028	-0.090	0.299
REGION6	0.097***	0.031	0.605**	0.243
REGION7	0.037	0.027	0.279	0.247
City	-0.007	0.017	-0.035	0.105
KIDS	-0.009	0.020	-0.058	0.130
Lagged Unemployed	0.324***	0.048	1.089***	0.141
SRH	-0.030**	0.013	-0.208***	0.075
Lagged SRH	0.022*	0.013	0.135*	0.078
Constant	0.459	0.284	0.363	1.353
N	1294		N	1294
F 21, 1272	6.82		Wald chi221	140.62
Prob > F	0.0000		Prob > chi2	0.0000
R-squared	0.1717		Pseudo R2	0.1969
Root MSE	.28603			

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Section 2.: Regression on differenced variables

In this section, I take differences of the SRH in order to treat simultaneity. By taking differences, I attempt to control for the endogeneity inherent in the model, as discussed in Chapter 3. I present two different models. First, I regress ΔSRH on the usual set of individual control variables and a set of binary variables for each of the possible transitions (or non-transitions) between work and unemployment (I use "transition" from work to work as my reference-group). Later, I use the unemployment variable as the dependent variable and add ΔSRH along with the usual set of regressors in order to see how changes in health status will influence the probability of being unemployed.

ΔSRH has an approximate normal distribution with values within the range of -4 and 4. It has a slight positive disparity for females. A positive value indicates an improvement in SRH, while a negative value indicates deterioration within the two surveys.

Table 4.: Health Equation – Estimation with Differences

	OLS		Ordered probit	
Dsrh1	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
AGE	0.071	0.289	0.063	0.375
AGE2	-0.039	0.035	-0.047	0.046
BMI3	-0.096	0.053	-0.126*	0.069
BMI4	-0.117*	0.065	-0.157*	0.085
Smoking	-0.091*	0.049	-0.118*	0.064
Subj. Wealth	-0.016*	0.015	-0.022	0.02
Female	0.039	0.047	0.050	0.061
TRANS1	-0.227**	0.097	-0.297**	0.124
TRANS2	-0.350***	0.134	-0.454***	0.170
TRANS4	-0.302***	0.113	-0.397***	0.147
Married	-0.024	0.062	-0.036	0.081
City	-0.071	0.046	-0.092	0.060
Kids	0.007	0.054	0.017	0.069
Constant	1.016*	.551		
N	1295		N	1295
F(13, 1281)	9.64		Wald chi2(13)	116.18
Prob > F	0.0000		Prob > chi2	0.000
R-squared	0.0869		Pseudo R2	0.037
Root MSE	.82337			

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Clearly, the differencing of the SRH led to a general loss in the significance of the variables. By only analyzing the change in health status, I restrict my focus to short and medium-term effects, to those that will have a significant effect on health within three years. With the exception of the transition indicators, age and to a limited extent BMI and smoking, all other marginal effects are insignificant in Table 20 in the Appendix.

All three transition variables indicate a negative correlation between transitions and SRH (when compared TRANS3 (work → work)), and their size seems to indicate that the length of unemployment also plays a role in health deterioration. TRANS4 (unemployment → unemployment) represents long-term unemployment; it is not surprising therefore that this variable has the highest marginal effect. Remaining unemployed throughout the two surveys adds to the probability of having experienced a large deterioration in SRH (ΔSRH $\leq$-2) by up to 9.2%. Surprisingly, TRANS4 (unemployment→ work) produced the second largest marginal effects on health and not TRANS1 (work → unemployment) as I expected. The most likely explanation for this is that those in TRANS4 have been in unemployment for longer than those in TRANS1, hence their deterioration in health will also be greater.[10] TRANS2 increases the probability of having experienced no change in SRH (ΔSRH =0) by 12%, i.e. if someone transitioned from unemployment to work, s/he was 12% less likely to have indicated the same SRH during both surveys.

The effects of AGE and AGE2 seem insignificant individually, however their joint significance is very high (F-statistic= 47.6). The marginal effect is the largest around ΔSRH=0, that is the probability of experiencing no change (or small change) in SRH within the three years will increase by 0.4% for each additional year of age. The marginal effect of age on health at large negative changes in SRH gets progressively smaller, while the marginal effect at positive changes in SRH is approximately zero.

Marginal effects of the differenced unemployment equation are presented in Table 21. of the Appendix. My conclusions regarding age are quite different in this specification. AGE and AGE2 are only moderately significant (F-statistic: 7.7) One additional year of age will lead to a

[10] There is a survey question regarding the length of unemployment in the 2005 survey, but it only has information for those who were unemployed in that year, ie. TRANS1 and TRANS2.

4.6% decline in the probability of unemployment, with the lowest probability of unemployment reached at the maximum of 55 years. Education remains the main indicator of unemployment. Each additional level of education attained will lead to around 40% decrease in the probability of unemployment.[11] Changes in SRH are only significant at 5%, but the size of the marginal effects are quite large. A unit increase in SRH will lead to 22.6% decrease in the probability of unemployment.

Overall, my results show that age, employment transition, and to a certain extent BMI and smoking will determine the change in SRH, while education, ΔSRH and regional differences are the main determinants of the probability of unemployment, which is in line with my findings in Section 1.

[11] Comparing the decline in probability of unemployment due to one more level of education attained in this model (≈40%), with that obtained in Section 1 (≈10%) , casts doubt on the accuracy of this model. Although the estimation in Section 1 suffers from endogeneity bias, the difference between the two estimations should not be so larger.

Table 5. Unemployment Equation – Estimation with Differences

	OLS Model		Probit Model	
Unempl.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
AGE	-0.048	0.117	-0.234	0.609
AGE2	0.002	0.014	0.008	0.075
EDU2	-0.245	0.18	-0.694	0.48
EDU3	-0.347*	0.178	-1.066**	0.477
EDU4	-0.389**	0.179	-1.341***	0.492
EDU5	-0.387**	0.179	-1.338***	0.497
EDU6	-0.414**	0.178	-1.733***	0.52
REGION2	0.019	0.028	0.166	0.256
REGION3	0.059*	0.032	0.399	0.251
REGION4	0.033	0.029	0.263	0.251
REGION5	-0.018	0.028	-0.092	0.293
REGION6	0.113***	0.034	0.661***	0.243
REGION7	0.043	0.029	0.313	0.246
BMI3	-0.006	0.019	-0.002	0.117
BMI4	-0.007	0.023	-0.017	0.141
Smoking	0.037**	0.019	0.201*	0.103
Subj. Wealth	-0.014	0.006	-0.080**	0.032
Female	-0.007	0.018	-0.014	0.106
Married	-0.028	0.022	-0.126	0.126
City	-0.017	0.017	-0.101	0.104
Kids	-0.004	,020	-0.05	0.125
ΔSRH2	-0.176	0.103	-0.684**	0.322
ΔSRH3	-0.195*	0.1	-0.806	0.301
ΔSRH4	-0.226**	0.1	-0.97***	0.306
ΔSRH5	-0.259**	0.103	-1.224***	0.369
Constant	0.872***	0.206		
N	1294		N	1294
F(25, 1268)	3.89		Wald chi2(25)	88.52
Prob > F	0		Prob > chi2	0
R-squared	0.1001		Pseudo R2	0.1365
Root MSE	0.29861			

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Section 3: IV Regression

In the following section, I attempt to use instrumental variable regression in order to fix the endogeneity problem. IV regression involves substituting the endogenous variable with an appropriate instrument, which is strongly correlated with the endogenous variable but is not correlated with the error term. In the context of my analysis this will look like the following: as before I use SRH and unemployment as the two dependent variables and add the entire set of individual characteristics variables used in the previous model as regressors. However, I replace the endogenous variables with valid instruments. A valid instrument must satisfy two conditions: It must be

1. Strongly correlated with the endogenous variable, but
2. Uncorrelated with the error term

Using the IV-method is an effective technique for treating endogeneity. However, if the instrument does not meet the above two assumptions it can produce dubious results. I was not able to find valid instruments for unemployment that met both requirements.[12] After analyzing my results, I discuss possible shortcomings.[13]

In the previous sections I have concentrated on analyzing the marginal effects of the non-linear estimations. Unfortunately in this section I am restricted to using a linear model, therefore the results in this section and in the previous sections will not be directly comparable.

Instrumenting Unemployment

Finding a valid instrument for unemployment is a difficult task. Most variables in my dataset are SES type variables and by definition, they are correlated with health through many channels. Below I use *education* and *regions* as my instruments. In order to prove that these are valid IVs I must prove that they only affect health through their correlation with unemployment. It is often argued that one's level of education and region of residence will affect health (through differences in local average income and better access to healthcare etc.), however this relationship does not seem significant in my specification. Education and region were insignificant in the lagged model in Section 1. On the other hand, the two variables could influence health through other channels (eg. through their correlation with family background), which I did not control for in my model.

[12] In addition to education and regions, I also tried regional unemployment, several other SES, mental and life-perception indices as possible instrumental variables. The conclusion in all cases was similar to what is presented for education and regions. The validity of the instrument is questionable either due to Condition 1. or Condition 2.
[13] One obvious choice for the instrumental variable could be the lagged version of unemployment. We know that it must be exogenous, since it is one time period removed from the present, thus it cannot be correlated with the current period error term. However, lagged unemployment turned out significant in Table 3, therefore it must be included in the IV-regression as an exogenous regressor. Leaving it out and using it as an instrument would lead a bias in the coefficients.

Below, I introduce two tests (the Hausman Endogeneity Test and the Test of Weak IV) in order to verify the above conditions and the specification of the model:[14]

Test of endogeneity

It is not possible to test the endogeneity of a variable *ex ante*, however under the assumption that a valid instrument has been found one can test the endogeneity of the regression by comparing OLS and IV-regression results using the Hausman-test. If the null-hypothesis of exogeneity is rejected, then the variable initially perceived as endogenous was indeed endogenous and using an IV-regression is justified.

Table 6: Hausman Test of Endogeneity:

	Chi-squared statistic	P-value
Regions and Education	0.145	0.704
Regions	2.488	0.115
Education	2.127	0.145

As seen in Table 6, the null-hypothesis of exogeneity cannot be rejected in either of the specifications. This indicates that either (1) endogeneity is not present in the regression, which would mean that OLS is the most appropriate method of estimation or (2) the instruments are not valid and they have the same direction of bias as the original endogenous variable. In Section 1, I have already established that simultaneity is present in the model (the two endogenous variables were both significant at "t"), therefore it is more likely that (2) holds and my instruments suffer from the same bias as the endogenous variable.

Test of Weak IV

Using an instrument that is only weakly correlated with the endogenous variable could add more bias to the estimation. In order to see whether Condition 1 holds, I test the strength of the instrument using the Cragg-Donald Test. This test is the equivalent to performing an F-test on the excluded instrument in the first-stage regression. According to the general rule of thumb the F-statistic should be above 10 to guarantee a strong enough IV. According to this test both

[14] In addition to the two test presented above, the test of overidentification should be performed if the number of instruments exceeds the number of endogenous variables.

education and *region* are weak IVs, their joint-significance and individual measured by the F-statistic is well below 10.

Table 7: Cragg-Donald Test of weak IV

	F-statistic
Regions and Education	6.01
Regions	5.95
Education	7.66

Analysis

According to the Hausman-test education and region may not be a valid IV as per Condition 1 and 2, however it is still worth evaluating the coefficients using the IV-regression. I chose *education* as my only instrument, since it proved to be a stronger IV than *education* and *region* combined. IV regression output with *education* is presented in Table 8.

Table 8.: IV regression with education

First Stage Regression			**IV regression**		
Unemployment	Coef.	Robust Std. Err.	SRH	Coef.	Robust Std. Err.
AGE	0.009**	0.010	**Unemployment**	**-0.659***	**0.344**
AGE2	0.000***	0.000	AGE	-0.024	0.023
BMI4	-0.013**	0.016	AGE2	0.000	0.000
BMI3	-0.016**	0.019	BMI4	-0.102***	0.038
Smoking	0.009**	0.015	BMI3	-0.245***	0.048
Subj. Wealth	-0.012***	0.005	Smoking	-0.119***	0.036
Female	-0.024**	0.015	Subj. Wealth	0.027**	0.013
Married	-0.015**	0.019	Female	-0.068*	0.036
Kids	-0.004**	0.017	Married	-0.018	0.046
Lagged Unemployed	0.334**	0.043	Kids	0.006	0.040
Lagged SRH	-0.001**	0.010	Lagged Unemployed	0.003	0.140
EDU1	0.375	0.191	Lagged SRH	0.277***	0.026
EDU2	0.162**	0.029	Constant	3.777***	0.449
EDU3	0.044**	0.017			
EDU4	0.020**	0.015			
EDU5	0.012**	0.019			
Constant	-0.072	0.211			
N	1590		N	1590	
F(16. 1573)	9.710		F(12. 1577)	28.65	
Prob > F	0.000		Prob > F	0.00	
Centered R2	0.170		Centered R2	0.17	
Uncentered R2	0.250		Uncentered R2	0.97	
Root MSE	0.270		Root MSE	0.66	

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

The coefficient on unemployment is negative and significant at 10% confidence level. However, when compared with results in Table 3. (in Section 1.) it is considerably larger than what I expected. The IV regression results indicate that unemployed individuals have SRH measure that is on average 0.66 units lower than working individuals', compared to the 0.16 units difference when OLS is used.[15] The standard errors on unemployment are also very large, which suggests a weak coefficient estimate. As noted in Section 3 my expectation is that the coefficients will decrease in absolute value after simultaneity is controlled for. This model produces significantly *larger coefficients*, which also leads to suspicion. The difference in coefficients could be due to a combination of factors that are not explicitly controlled for in the regression but are somewhat correlated with education and health, for example the rate of local employment, proximity to major healthcare centers, and general infrastructure. Individually these factors would be insignificant in the regression, but due to their correlation with education they will influence health. Then again, education cannot be a valid instrument if its influence on SRH is not strictly through unemployment, but through a series of other factors as well.

In general, the OLS model in Table 3 and the IV-model above produce similar coefficients with the exception of lagged unemployment and the female variable. AGE and AGE2 seem insignificant but when tested jointly they appear highly significant. BMI remained significant while the relevance of subjective wealth has decreased.[16] In terms of coefficient size, the instrument has the largest effect (-0.66) in absolute value, followed by BMI2 (-0.24) and the smoking variable (0.11). Overall, due the Hausman-test and the unexpectedly large coefficient on unemployment I conclude that this IV regression did not produce reliable result.

[15] The regression coefficients in Table 3. are biased upwards in absolute value by simultaneity. Therefore, if simultaneity is controlled for, they should decline.

[16] It should be noted here that there is likely some degree of simultaneity between BMI-index and SRH. Being overweight could lead to illness, but on the flipside having a low body height (and subsequently high BMI-index) could indicate nutritional deficiency during early childhood, which would likely lead to lower health status during one's adult life. However, when I instrument BMI-index using lagged values (a valid IV for BMI-index), the Hausman-test fails to reject full exogeneity. Therefore, the instrumentation of BMI-index in my specification of the model is not necessary.

Instrumenting Health

I use BMI and smoking variables as my instruments for health. Again, in order to prove that this is a valid IV, I must prove that it is strongly correlated with SRH, but that it is only correlated with unemployment through health. Proving the former requirement is straightforward: both BMI and smoking are highly significant in the first-stage regression as seen in Table 12 However, the latter may be more difficult. Body weight and smoking habit may be independent of health in the short-term. Their detrimental effects on health may only be realized if obesity and smoking habits are maintained in the long-run. Therefore, one could argue that there is a range in which the instruments are independent of health, but they are correlated with unemployment. For example, one could claim that there is a positive correlation between psychological stress and the probability of unemployment, and since stressful lifestyle may lead to smoking and weight gain they too will display positive correlation. This would mean that the instruments are correlated with unemployment through stress without the significant deterioration in health.

Test of endogeneity

I use the Hausman Test of Endogeneity as before in order to prove that SRH is indeed endogenous. I present the test results for the two variables individually and jointly in Table 9.

Table 9: Hausman Test of Endogeneity:

	Chi-squared statistic	P-value
BMI	0.233	0.629
Smoking	1.962	0.161
BMI and Smoking	0.560	0.454

Strict exogeneity cannot be rejected according to either of the specifications. If I assume that the IVs are valid, this would indicate that SRH is exogenous and there is no need to treat endogeneity. However, since simultaneity has been established in Section.1, it is more likely that either simultaneity is not strong enough to result in considerable change in the coefficients for the Hausman-test to detect it, or the IVs suffer from the same bias as the endogenous variable.

Test of weak IV

All three specifications have an F-statistic above 10. Hence, they are strongly correlated with the SRH, as per the Test of Weak IV (explained above).

Table 10.: Cragg-Donald Test of weak IV

	F-statistic
BMI	24
Smoking	26.6
BMI and Smoking	24

Test of Overidentifying Restrictions

If the number of instruments exceeds the number of endogenous variables, we can test whether the additional instruments are valid by Condition 2. The joint null hypothesis is that the instruments are valid, and that the excluded instruments are correctly specified. If the null hypothesis is rejected, then one or more of the instruments are invalid. I use Hansen J-test to test overidentification. This test cannot be performed for the instruments individually, since in that case the estimation is just-indentified (the number of instruments equals the number of regressors).[17] The Hansen J-test cannot reject the validity of the instruments. Therefore, according to this statistic, I can accept them as valid.

Table 11.: Hansen J test

	Chi-squared statistic	P-value
BMI and Smoking	1.585	0.454

Analysis

[17] Although technically the Hansen J-test could be performed for BMI alone (as it consists of two variables BMI3 and BMI4), this is conceptually irrelevant since I would be testing the validity of one BMI variable by comparing it to the other.

Although the Hausman-test indicates that endogeneity is not present in the model, it is worth evaluating the output. First of all, it is worth noting that the sign and the size of the SRH coefficient in the IV-regression (-0.065) is correct. As noted before, I expect a negative correlation between SRH and unemployment, since someone with lower SRH is more likely to become unemployed. Also, I expect the sign of the coefficient after treating simultaneity to be smaller (or at least in the same range) as it was before simultaneity was treated. Although 0.065 is slightly larger than the coefficient (0.030) in Section 1 it is in the same range, therefore the strength of the instrument and its general direction should be correct. However, SRH is not significant anymore (with the p-value of 0.457).

There could be at least two reasons for this result:

(1) it could be the case that the negative correlation between poor health and unemployment is not present after simultaneity is treated, or

(2) the relationship could be non-linear, which would result in an insignificant coefficient in the linear IV-regression model.

At this point, there is no way to answer this question precisely. However, the likely solution will become clearer in the next section, when the simultaneous non-linear model is analyzed.

Table 12. IV Regression with BMI and Smoking

First Stage Regression			IV regression		
SRH	Coef.	Robust Std. Err.	Unempl.	Coef.	Robust Std. Err.
AGE	-0.022	0.232	**SRH**	**-0.065**	**0.088**
AGE2	-0.02	0.029	AGE	0.011	0.116
EDU2	-0.142	0.199	AGE2	-0.005	0.014
EDU3	-0.09	0.196	EDU2	-0.193	0.172
EDU4	-0.02	0.198	EDU3	-0.271	0.171
EDU5	-0.017	0.202	EDU4	-0.309*	0.172
EDU6	0.01	0.199	EDU5	-0.316*	0.172
Subj. Wealth	0.023**	0.013	EDU6	-0.323	0.172
Female	-0.058	0.040	Subj. Wealth	-0.010	0.006
Married	-0.018	0.049	Female	-0.007	0.017
REGION2	-0.070	0.085	Married	-0.024	0.022
REGION3	-0.062	0.090	REGION2	0.018	0.028
REGION4	-0.007	0.082	REGION3	0.042	0.031
REGION5	-0.030	0.084	REGION4	0.036	0.028
REGION6	0.055	0.084	REGION5	-0.017	0.028
REGION7	-0.067	0.083	REGION6	0.099***	0.032
City	-0.032	0.040	REGION7	0.035	0.028
Kids	-0.015	0.045	City	-0.008	0.017
Lagged Unempl.	-0.234***	0.068	Kids	-0.009	0.02
Lagged SRH	0.283***	0.030	Lagged Unempl.	0.315***	0.052
BMI3	-0.109**	0.042	Lagged SRH	0.032	0.029
BMI4	-0.190***	0.052	Constant	0.576	0.41
Smoking	-0.148***	0.040			
Constant	3.399***	0.477			
N	1295		N	1295	
F(12, 1282)	23.92		F(10, 1284)	9.010	
Prob > F	0.000		Prob > F	0.000	
Centered R2	0.202		Centered R2	0.122	
Uncentered R2	0.968		Uncentered R2	0.217	
Root MSE	0.661		Root MSE	0.292	

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Section 4 : Simultaneous Regression

In the following section, I estimate the system in a simultaneous fashion in order to perfectly control for the simultaneity between the two endogenous variables. By estimating the two equations in parallel, I permit the interaction of the two endogenous variables in the same time period. Below I present the simultaneous Three-Stage Least Squares (3SLS) version of the model. 3SLS is simply another IV-regression in the sense that each equation (health and unemployment) individually serve as the first stage regressions of the dependent variable. In other words, in the first and second stage I run a regression on the health-equation and on the unemployment-equation to instrument SRH and unemployment, respectively. In the third stage, the two instrumented equations are used in place of the endogenous variable in the other equation. As before, the IVs must be valid in order to obtain appropriate estimates. The instruments I use will be the same as in Section 3: education is used to instrument unemployment, and BMI and smoking utilized to instrument health.

In Table 13. I present a linear 3SLS estimate, and a non-linear estimation using the *bioprobit* command in STATA. *Bioprobit* fits a maximum-likelihood two-equation ordered probit models of ordinal variables.[18] [19]The significance of the linear and non-linear models are very different. The two endogenous variables are not significant in the linear regression, while they are highly significant in the non-linear model. (Reminder: unemployment was weakly significant in the Health IV-Model, while SRH was insignificant in the Unemployment IV-Model). This suggests that the relationship between the two dependent variables could be non-linear, which would also explain why the endogenous variable coefficients in linear one-way IV estimations were insignificant in Section 3. Comparing the linear coefficient sizes in Table 13 above with those in Table 3 in Section 1, we see that all variables except the two endogenous ones remained within the same range. Unemployment in the health-equation increased in absolute terms from -0.16 to -0.32, while SRH in the unemployment-equation -0.06 to -0.03 in absolute terms. Interestingly, while the significance of the endogenous variables have increased, the significance of other explanatory variables, such as smoking and subjective wealth in the health-equation and education and regions in the unemployment-equation, have decreased. In fact, after controlling for simultaneity subjective wealth seems statistically insignificant. (Reminder: subjective wealth was weakly significant in Section 1) If my specification is correct, this would indicate that one's

[18] In this estimation one dependent variable (unemployment) is binary, while the other (SRH) is ordered type. Ideally, for this estimation I should run a probit/ordered-probit maximum likelihood regression, but unfortunately, this command does not exist.

[19] Unfortunately, the bioprobit command does not allow any postestimations of regression. For this reason and due to the lack of time I do not analyze the marginal effects of this model. The linear and non-linear model coefficients cannot be directly compared in Table 13, therefore I will restrict my analysis to non-linear coefficient significance and I will disregard coefficient size.

subjective perception of wealth does not contribute to differences in health status, after treating endogeneity. Even more intriguing is the lack of significance on the education binary variables in Table 13. (Reminder: In Section 1, education is the main predictors of unemployment according to the non-linear estimates in Table 3 and according to the marginal effects in Table 19. An additional level of education completed decreases the probability of unemployment by approx. 10%) This could be a result of the correlation of the BMI and smoking variables with education. In fact, adding BMI and smoking to the probit estimation in Section 1, leads to a minor decrease in the coefficients and in the significance of the education variables.[20]

Overall, the simultaneous model indicates that unemployment will lead to a significant deterioration in health and that poor health will lead to a significant increase in the probability of unemployment after simultaneity problem has been treated. However after controlling for endogeneity the significance of other explanatory variables, such as subjective health and education, will also decrease. According to the non-linear specification only the endogenous variables, the lagged versions of the dependent variables and BMI have any significance.

[20] If this explanation is true, then BMI and smoking cannot be valid instruments for health, since they will influence unemployment through other means than health. BMI and smoking are negatively correlated with education independent of health status, which is negatively correlated with unemployment.

Table 13.: Linear and Non-linear 3SLS Regression

	Linear Model		Bioprobit	
	Coef.	Std. Err.	Coef.	Std. Err.
SRH				
AGE	-0.032	0.227	-0.103	0.378
AGE2	-0.023	0.028	-0.03	0.047
BMI3	-0.105**	0.042	-0.061**	0.030
BMI4	-0.194***	0.051	-0.114**	0.060
Smoking	-0.158***	0.040	-0.087	0.054
Subj. Wealth	0.029**	0.0132	0.032	0.020
Female	-0.052	0.038	-0.061	0.063
Married	-0.022	0.050	-0.062	0.083
City	-0.024	0.038	-0.062	0.062
Kids	-0.013	0.050	-0.027	0.075
Unempl.	**-0.319**	**0.320**	**-1.513*** **	**0.095**
Lagged Unempl.	-0.129	0.133	0.133	0.118
Lagged SRH	0.288***	0.026	0.464***	0.045
Constant	3.354***	0.450		
Unempl.				
AGE	0.001	0.099	-0.156	0.383
AGE2	-0.005	0.012	-0.023	0.047
EDU2	-0.173*	0.102	-0.067	0.086
EDU3	-0.253***	0.101	-0.101	0.080
EDU4	-0.297***	0.101	-0.122	0.095
EDU5	-0.305***	0.102	-0.129	0.103
EDU6	-0.314***	0.102	-0.139	0.101
Subj. Wealth	-0.010	0.006	0.027	0.025
Female	-0.007	0.017	-0.038	0.067
Married	-0.023	0.021	-0.058	0.104
REGION2	0.0243	0.033	0.045	0.081
REGION3	0.0468	0.033	0.047	0.114
REGION4	0.036	0.032	0.054	0.112
REGION5	-0.014	0.036	0.014	0.108
REGION6	0.091***	0.033	0.080	0.100
REGION7	0.040	0.0315	0.048	0.131
City	-0.006	0.017	-0.070	0.063
KIDS	-0.008	0.019	-0.040	0.080
Lagged Unemployed	0.316***	0.037	0.215*	0.124
SRH	**-0.060**	**0.087**	**-1.47*** **	**0.053**
Lagged SRH	0.030	0.028	0.451***	0.050
Constant	0.541	0.360	15.010	450.600
N	1294		N	1294
Prob>chi2	0.000		Wald chi2(13)	550.43
			Prob>chi2	0.000

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Conclusion

The purpose of this study was to analyze the causal relationship between unemployment and health, and to examine whether any significant relationship exist after simultaneity is treated. After establishing the presence of simultaneity, I use three different methods to circumvent the problem. The conclusion throughout the four models is consistent. Unemployment has a detrimental effect on health, and vice versa poor health will lead to a higher probability of unemployment after endogeneity is treated.

In Section 1, I find that current and past unemployment as well as past health-status are strongly significant in determining current health, and conversely that current health-status and past unemployment will determine unemployment. Using this model I also establish the presence of simultaneity and causality leading from unemployment towards health. By analyzing the marginal effects of the non-linear health estimation, I find that in addition to unemployment, age, BMI and smoking are the major determinants of health, and that despite expectations the contribution of subjective wealth is only limited. On the other hand, looking at the marginal effects of the unemployment-equation I find that in addition to a history of unemployment age, education and region are the main determinants of the probability of becoming unemployed.

Similar results appear from Section 2, where I evaluate how employment transitions affect the change in health-status between the two survey waves. I find that the deterioration in health is the greatest for those who were unemployed during both surveys. Less health deterioration can be measured among those who transitioned from unemployment to work during the two surveys, and the least amount of deterioration is noticeable among those who transitioned from work to unemployment. Examining the marginal effects reveals that employment transitions are the most significant in explaining a change in health if the change in health-status is moderate ($-1<\Delta SRH<1$), which means that transitions will not lead to extreme changes in health.

In Sections 3 and 4, I use two different methods of instrumental variable regression to fix the endogeneity problem. Assuming that my choice of instruments is valid, I can conclude that the two endogenous variables remain significant after simultaneity is treated using a simultaneous regression, however that the significance of all other explanatory variables with the exception of BMI, such as education and region disappear.

Reference:

Ahs, et. al. 2006 Self-rated health in relation to employment status during periods of high and of low levels of unemployment, The European Journal of Public Health 2006 16(3):294-304;

Antal, Z.L, 2005: A „kisebbségi" és a „többségi" betegutak sajátos vonásai, Társadalomkutatás, Volume 22, Numbers 2-3 / August 2004, p. 335-366

Blakely, T. A., Lochner, K. and Kawachi, I. (2002) Metropolitan area income inequality and self-rated health – A multi level study. Social Science & Medicine, 54, 65-77

Böckerman, Petri & Ilmakunnas, Pekka, 2007. "Unemployment and self-assessed health: Evidence from panel data," MPRA Paper 1798, University Library of Munich, Germany.

Cooper ,D, W. David McCausland and Ioannis Theodossiou, 2007b, Unemployed, Uneducated and Sick: The Effects of Socioeconomic Status on Health Duration in the European Union, University of Aberdeen Business School Working Paper No. 2007-11

Cooper, D , W. David McCausland and Ioannis Theodossiou, 2006, The Health Hazards of Unemployment and Poor Education: the Socioeconomic Determinants of Health Duration in the European Union, Economics & Human Biology, Vol. 4, No. 3, pp. 273-297, December 2006

Cooper, D, McCausland, W.D, Theodossiou, I, 2007a: The Health Hazards of Unemployment and Poor Education: The Socioeconomic Determinants of Health Duration in the European Union, March 2007 University of Aberdeen Business School Working Paper

Dahl, Espen, 1993 "Social Inequality in Health – the role of the healthy worker effect" Soc Sci Med, 36:1077 – 1086

Firle Réka, Scharle Ágota és Szabó Péter András (2008): A rendszeres szociális segély munkakínálati hatása, Munkaerőpiaci Tükör 2008,p. 75-87.

Gatignon, Hubert: Statistical Analysis of Management Data, Springer, 2003

Gash, Vanessa & Mertens, Antje & Romeu Gordo, Laura, 2006. "Are fixed-term jobs bad for your health?: a comparison of West-Germany and Spain," IAB Discussion Paper 200608, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

Gerdtham, U. –G. & Johannesson, M., 2003. "A note on the effect of unemployment on mortality" Journal of health Economics, 22:5005-518

Gerdtham, U. –G. & Johannesson, M., 2001 "The relationship between happiness, health and socioeconomic factors: results based on Swedish microdata" Journal of Socioeconomics, vol. 30, pp. 553 – 57

Goldsmith, Arthur H. & Veum, Jonathan R. & William Darity, Jr., 1996. "The impact of labor force history on self-esteem and its component parts, anxiety, alienation and depression," Journal of Economic Psychology, Elsevier, vol. 17(2), pages 183-220, April.

Gordo, Laura Romeu 2006. "Effects of short- and long-term unemployment on health satisfaction: evidence from German data," Applied Economics, Taylor and Francis Journals, vol. 38(20), pages 2335-2350, November.

Gyukits György (2000): A romák egészségügyi ellátásának szociális háttere. In: Cigánynak születni, Aktív Társadalom Alapítvány, Új Mandátum Kiadó, Budapest.

Jin, Robert L. et al (1995) The impact of unemployment on health: a review of the evidence, Canadian Medical Association Journal, Sept. 1, 1995 153(5)

Huisman, et al: The predictive ability of self-assessed health for mortality in different educational groups, International Journal of Epidemiology, 2007:36:1207-1213

Kopp, M. S., Csoboth, Cs., & Réthelyi, J. (2004A). Psychosocial determinants of premature health deterioration in a changing society: The case of Hungary. Journal of Health Psychology, 9, 99–109.

Kopp, M.S., & Skrabski, A., Kawachi, I., & Adler, N. (2004B) Low socioeconomic status of the opposite sex is a risk factor for middle aged mortality. J. Epidemiol. Community Health 2005;59;675-678

Kuhn, A, & Rafael Lalive & Josef Zweimüller, 2007. "The Public Health Costs of Unemployment," Cahiers de Recherches Economiques du Departement d'Econometrie et d'Economie politique (DEEP) 07.08, Université de Lausanne, Faculté des HEC, DEEP.

Moffitt, , R. A. (2002): Welfare programs and labor supply, NBER Working Papers No. 9168.

Moser KA, Fox AJ, Jones DR. Unemployment and mortality in the OPCS Longitudinal Study. *Lancet.* 1984 Dec 8;2(8415):1324–1329

Sinn, Hans-Werner Redesigning the welfare state: Germany's current agenda for an activating social assistance, Edward Elgar Publishing, 2006

Simon et al, How is your health in general? A qualitative study on self-assessed health, European Journal of Public Health, Vol. 15. 2, 200-208 2005

Tahin Tamás–Jeges Sára–Lampek Kinga (2000): Az iskolai végzettség és az egészségi állapot változása követéses vizsgálat alapján. Demográfia, 43.p. 305–334.

Takács Erika, Szaszkó Dóra, Dr. Belicza Éva, Dr. Boncz Imre 2006 Az OEP járóbeteg szakellátási kassza igénybevételének területi egyenlőtlenségei

Appendix:

Table 14: SRH and age

	Wave 1.			Wave 2.		
SRH	Freq.	Age	St. Dev.	Freq.	Age	St. Dev.
1	17	39.1	10.4	15	47.6	7.2
2	166	39.1	9.5	79	46	9.1
3	1088	38.1	9.6	667	44.5	9.1
4	459	40.1	10	907	40.4	9.6
5	84	38.5	11	147	36.5	8.8
Total	1814	38.7	9.8	1815	41.9	9.7

Table 15: Age-groups and Unemployment

	Wave 1.			**Wave 2.**		
	Working	Unempl.	Freq.	Working	Unempl.	Freq.
25-35	92%	8%	543	89%	11%	490
36-45	93%	7%	561	90%	10%	541
46-55	94%	6%	518	91%	9%	614
Total	92%	8%	1,815	90%	10%	1,815

Table 16. Education and Unemployment

	Wave 1.			Wave 2.		
	Working	Unempl.	Freq.	Working	Unempl.	Freq.
Less than 8 years of education	75%	25%	16	44%	56%	9
8 years of education	82%	18%	273	75%	25%	237
Vocational training without high school diploma	92%	8%	611	88%	12%	632
Vocational training with high school diploma	95%	6%	400	93%	7%	386
High school diploma	93%	7%	222	94%	6%	231
College or university diploma	99%	1%	288	98%	2%	319
Total	92%	8%	1810	90%	10%	1814

Table 17 : Distribution according regions

	Frequency	Percentage
Central Hungary	249	13.72
Southern Transdanubia	256	14.10
Northern Great Plain	218	12.01
Western Transdanubia	278	15.32
Central Transdanubia	222	12.23
Northern Great Plain	288	15.87
Northern Hungary	304	16.75
Total	1,815	100

Table 18: Marginal effects in the health-equation using ordered probit-estimation

	SRH=1		SRH=2		SRH=3		SRH=4		SRH=5	
	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.
AGE	0.001	0.0029	0.006	0.026	0.029	0.123	-0.026	0.109	-0.010	0.043
AGE2	0.000	0.00036	0.002	0.003	0.011	0.015	-0.009	0.013	-0.004	0.005
BMI4	0.001*	0.0008	0.013**	0.005	0.061***	0.022	-0.055***	0.021	-0.021***	0.008
BMI3	0.003***	0.00165	0.027***	0.009	0.105***	0.026	-0.103***	0.028	-0.032***	0.007
Smoking	0.002**	0.00097	0.020***	0.005	0.085***	0.021	-0.079***	0.020	-0.029***	0.007
Subj. Wealth	-0.000*	0.00022	-0.003**	0.001	-0.016**	0.007	0.015**	0.006	0.006***	0.002
Female	0.001	0.0005	0.006	0.004	0.030	0.020	-0.026	0.018	-0.010	0.007
Married	0.000	0.00057	0.002	0.005	0.010	0.026	-0.009	0.023	-0.004	0.009
City	0.000	0.00047	0.002	0.004	0.008	0.020	-0.007	0.018	-0.003	0.007
Kids	0.000	0.00055	0.001	0.005	0.006**	0.023	0.005	0.021	-0.002	0.008
Unemployed	0.000*	0.0017	0.022**	0.011	0.084***	0.031	-0.084***	0.040	-0.026***	0.009
Lagged Unemployed	0.003*	0.00194	0.025**	0.012	0.093***	0.032	-0.094***	0.040	-0.028***	0.009
Lagged SRH	-0.003***	0.00132	-0.032***	0.005	-0.150***	0.018	0.133***	0.016	0.052***	0.007

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Table 19: Marginal effects in the unemployment-equation using probit-estimation

	Unemployed=1	
	Marg. Eff.	Std. Err.
AGE	-0.004	0.091
AGE2	-0.002	0.011
EDU2	-0.061	0.041
EDU3	-0.106*	0.058
EDU4	-0.107***	0.035
EDU5	-0.091***	0.021
EDU6	-0.114***	0.026
Subj. Wealth	-0.008*	0.005
Female	-0.005	0.015
Married	-0.022	0.020
REGION2	-0.026	0.043
REGION3	-0.052	0.050
REGION4	-0.048	0.050
REGION5	-0.012	0.038
REGION6	0.113**	0.056
REGION7	0.044	0.043
City	-0.005	0.015
KIDS	-0.008	0.018
Lagged Unemployed	0.265***	0.049
SRH	-0.029***	0.010
Lagged SRH	0.019*	0.011

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Table 20: Marginal effects of the differenced health-equation using ordered probit-estimation

	ΔSRH=-3		ΔSRH=-2		ΔSRH=-1		ΔSRH=0		ΔSRH=1		ΔSRH=2		ΔSRH=3		ΔSRH=4	
ΔSRH	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.
AGE	-0.002	0.013	-0.011	0.065	0.012	0.071	0.018	0.109	0.006	0.035	0.001	0.004	0.001	0.004	0.000	0.001
AGE2	0.002	0.002	0.008	0.008	0.009	0.009	-0.014	0.013	-0.004	0.004	-0.001	0.001	-0.000	0.001	-0.000	0.000
BMI4	0.004	0.003	0.022*	0.012	0.023*	0.012	-0.037*	0.020	-0.012*	0.006	-0.001*	0.001	-0.001*	0.001	-0.000	0.000
BMI3	0.006	0.004	0.028*	0.016	0.028**	0.014	-0.046*	0.025	0.014*	0.007	0.002*	0.001	-0.002*	0.001	-0.000	0.000
Smoking	0.004*	0.002	0.021*	0.012	0.022*	0.012	-0.035*	0.019	-0.011*	0.006	-0.001	0.001	-0.001	0.001	-0.000	0.000
Subj, Wealth	0.001	0.001	0.004	0.003	0.0.004	0.004	-0.006	0.006	-0.002	0.002	-0.000	0.000	-0.000	0.000	-0.000	0.000
Female	-0.002	0.002	-0.009	0.011	-0.010	0.012	0.015	0.018	0.005	0.006	0.001	0.001	0.001	0.001	0.000	0.000
TRANS1	0.014*	0.008	0.057**	0.026	0.044***	0.014	-0.089**	0.038	-0.023***	0.008	-0.002**	0.001	-0.002**	0.001	-0.000	0.000
TRANS2	0.025*	0.015	0.092**	0.038	0.054***	0.010	-0.136***	0.051	-0.031***	0.009	-0.003**	0.001	-0.003**	0.001	-0.000	0.000
TRANS4	0.020*	0.011	0.079**	0.033	0.051***	0.011	-0.120***	0.044	-0.028***	0.008	-0.003**	0.001	-0.003**	0.001	-0.000	0.000
Married	0.001	0.003	0.006	0.014	0.007	0.016	-0.010	0.023	-0.003	0.008	-0.000	0.001	-0.000	0.001	-0.000	0.000
City	0.003	0.002	0.016	0.010	0.017	0.012	-0.026	0.018	-0.009	0.006	-0.001	0.001	-0.001	0.001	-0.000	0.000
Kids	-0.001	0.002	-0.003	0.012	-0.003	0.013	0.005	0.020	0.002	0.007	0.000	0.001	0.000	0.001	-0.000	0.000

Note:* significant at 10%; ** significant at 5%; *** significant at 1%

Table 21: Marginal effects of the differenced unemployment-equation using probit-estimation

	Unemployed=1	
	Marg. Eff.	Std. Err.
AGE	-0.048	0.117
AGE2	0.002	0.014
EDU2	-0.245	0.180
EDU3	-0.347*	0.178
EDU4	-0.389**	0.179
EDU5	-0.387**	0.179
EDU6	-0.414**	0.179
REGION2	0.019	0.029
REGION3	0.059*	0.032
REGION4	0.0328	0.029
REGION5	-0.018	0.029
REGION6	0.113***	0.030
REGION7	0.043	0.029
BMI3	-0.006	0.020
BMI4	-0.006	0.023
Smoking	0.037*	0.019
Subj. Wealth	-0.014**	0.006
Female	-0.007	0.018
ΔSRH2 (=-1)	-0.176*	0.103
ΔSRH3 (=0)	-0.195*	0.099
ΔSRH4 (=1)	-0.226**	0.010
ΔSRH5 (>1)	-0.259**	0.103
Married	-0.028	0.022
City	-0.017	0.017
Kids	-0.004	0.021

Note:* significant at 10%; ** significant at 5%; *** significant at 1%