**Niels van Opstal 4-1**

**Introduction:**

The positive correlation between health and wealth has been shown many times. Pollack et al. (2007)⁠ for example found 29 studies testing for the correlation between health and wealth. But as they note in the discussion of their paper, “since most of the studies identified were cross-sectional, causal inferences cannot be made” (Pollack et al., 2007, p. 263). Health and wealth could easily be endogenous. Wealth could affect health through better access to health care and better living conditions. Health could affect wealth through the ability to work and thus accumulate wealth. It could also be possible that for example time preference affects both the accumulation of wealth and health (Meer, Miller, & Rosen, 2003)⁠.

As Aittomäki, Martikainen, Laaksonen, Lahelma, and Rahkonen (2010)⁠ point out, “Wealth in particular is not directly affected by changes in labour market participation that may radically alter the current income level”. This is important “… as changes in health and illness are likely to develop over a considerable time span” (Aittomäki, Martikainen, Laaksonen, Lahelma, & Rahkonen, 2010, p. 1025)⁠. They find that wealth is highly relevant in explaining health.

The goal of this paper is to examine the causality between health and wealth in the Netherlands. This paper will try to deal with the endogeneity between health and wealth using a difference in difference analysis. The treatment group consists of home-owners and the control group consists of people who rent a house. The treatment is the financial crisis in 2008 as it affected housing prices. The change in housing prices should only affect the wealth of the home-owners and it stands to reason that it does not affect or is affected by health status of the home-owner or home-renters alike. The wealth of a person will be defined as the net worth of a person, i.e. all their assets minus their liabilities. Each person will be placed into one of two categories, i.e. healthy (1) or not healthy (0), based on their self rated health.

To try to find the causal relationship between health and wealth the data from the DNB Household Survey (DHS) from CentERdata. This survey is described as “a unique data set allowing you to study both psychological and economic aspects of financial behavior. This panel survey was launched in 1993 and comprises information on work, pensions, housing, mortgages, income, possessions, loans, health, economic and psychological concepts, and personal characteristics. We have been collecting these data from 2,000 households participating in the CentERpanel” (“DHS data access | CentERdata.nl,” n.d.). Using different waves of this dataset will allow this paper to try and find a causal relationship using the difference in difference analysis.

The paper will be structured as follows. Chapter 2 will discuss relevant literature on the subject of causality between health and wealth. Chapter 3 will describe the data used and of what the variables consists of. Chapter 4 will describe the statistical methods used. Firstly there will be explained how there will be showed that there is a correlation in the data between health and wealth. Secondly the method that will be used to test for a causal relationship will be described. Chapter 5 will present the results found by the statistical analysis and chapter 6 will discuss the results and improvements for this study.

**Related literature:**

In their article *Should Health Studies Measure Wealth*, Pollack et al. (2007) systematically analyze a total of 29 articles that used health as the dependent variable and wealth and at least one other socioeconomic-status variable as independent variables. Of the 29 articles analyzed, 14 used self assed health as their health variable. Most of those articles reported positive or mixed results. The other 15 articles used different variables for health such as: mortality, chronic conditions, functional status and mental health. Of the total of 29 studies, 15 found positive results, 10 found mixed results and only 4 found negative results. They conclude that there is a significant correlation between healh and wealth. Especially when the wealth variables were constructed from detailed questions instead of simpler questions (for example just a single question). It should however be noted that they only check for correlation and do not address causality.

There are however some studies that do address causality in the health wealth connection and they mostly find negative results. Meer, Miller and Rosen (2003) use a straightforward instrumental variable strategy to deal with the endogeneity. They use inheritance as the instrument as it does affect health but does not directly affect health nor is it affected by health they reason. They do find a significant correlation between health and wealth but when inheritance is used as an instrument they do not find a significant effect from wealth on health. They conclude that short run changes in wealth do not affect health. They do however note: “This finding does not rule out the possibility of a long-term impact of wealth on health” (Meer, Miller, & Rosen, 2003, p. 729)⁠. Kim and Ruhm (2012)⁠ also use inheritance as exogenous wealth shocks and also find no significant effect on health.

In a similar study, Apouey and Clark (2015)⁠ also find small or negligible effects on general health using lottery winnings and inheritance as instruments. They do however find that lottery winnings do produce better mental health but also increase smoking and social drinking. They note that “health is not a holistic concept, and we need to both be clear about what kind of health we are talking about and be ready for the possibility that different types of health behave in very different ways” (p. 536). Au and Johnston (2015)⁠ even find that wealth shocks in the form of inheritance might even increase obesity in women.

Michaud and Soest (2008)⁠ also find no causal effects of wealth on health. They use a dynamic panel data model to test for the causality. As they note in their conclusion, the data they use consists only of elderly couples. They suggest that there might be a causal effect in different age groups and that it would be interesting to see if there are differences in different countries to see if institutions have an impact on the possible causal relationship.

There are some studies that do find a significant causal effect of wealth on health. Cai (2009)⁠ focuses on health transitions instead of health status itself to avoid the endogeneity of wealth and health. She finds that wealthy people are less likely of transitioning from healthy to unhealthy compared to people in the lower end of the wealth distribution. This, she argues, is evidence that there might be a causal effect of wealth on health. She proposes four different explanations of the causal effect of wealth on health. Firstly, because the study focuses on people in Australia malnutrition might not be an issue, eating less healthy food is associated with people with less economic recourses. Secondly, people with more wealth may live in better and healthier environments. Thirdly, even in a country with universal health care system such as Australia, wealthier people might still receive more health services that less wealthy people. Finally, wealth could give people more freedom in making decisions, thus experiencing less chronic stress which leads to poor health. So there are several ways in which wealth could exert an effect on health. Testing via which effect wealth does affect health was out of the scope of the paper.

**Data**

The data consists of different (yearly) waves from the DHS. The data is collected every year by the CentERdata. The DHS consists of six questionnaires, General Information on the Household, Household and Work, Accommodation and Mortgages, Health and Income, Assets and Liabilities and Economic and Psychological Concepts. (“DHS data access | CentERdata.nl,” n.d.)⁠

Besides the questionnaire data, the CentERdata also provides two aggregated data files, the aggregated income data and the aggregated wealth data. This paper will only use the Health and Income questionnaire which includes the self-rated health variable and the aggregated wealth data. The aggregated wealth data is made up from different questionnaires and consists of all the assets and liabilities someone might have.

The health variable will be a categorical variable with two options, either *healthy* or *not healthy*. The data received from the DHS has five categories for health: *poor, not so good, fair, good* and *excellent.* Persons who considered themselves to be in *poor* or *not so good* health will be places in the *not healthy* category. The persons that consider themselves to be in *fair, good* or *excellent* health will be placed in the healthy category. Self rated health is a good predictor for mortality (Idler & Benyamini, 1997)⁠ which is a good indicator for health.

To check if there is a correlation between wealth and health, a wealth variable needs to be created. For wealth this paper uses the net wealth of a person. Someone might well own a nice car and a house, but if he has a loan for the car and two mortgages on the house, he might still have a negative net wealth. To calculate the net wealth, all the assets of a person have been added together and the liabilities have been subtracted from the assets. The questionnaires are quite detailed on wealth which is important as was pointed out by Pollack et al. (2007).

The assets that could be found in the aggregated wealth data were:

Table Assets and liabilities of which net worth is made up off

|  |  |
| --- | --- |
| Assets | Liabilities |
| checking accounts | private loans |
| employer-sponsored savings plans | extended lines of credit |
| savings or deposit accounts | outstanding debts not mentioned earlier |
| deposit books | finance debts |
| savings certificates | loans from family or friends |
| single-premium annuity insurance policies | study loans |
| savings or endowment insurance policies | credit card debts |
| growth funds | loans not mentioned before |
| mutual funds and/or mutual fund accounts | checking accounts with negative balance |
| bonds and/or mortgage bonds |  |
| stocks and shares |  |
| put options bought |  |
| put options written |  |
| call options bought |  |
| call options written |  |
| pieces of real estate, not being used for own accommodation |  |
| value of life insurance mortgage real estate |  |
| cars |  |
| motorbikes |  |
| boats |  |
| (site-)caravans/trailers |  |
| money lent out to family or friends |  |
| savings or investments not mentioned before |  |
| stocks from substantial holding |  |
| business equity (professions) |  |
| business equity self-employed |  |
|  |  |



Figure Housing Prices in the Netherlands

This paper will use the waves from 2007, 2013 and 2017. As is clearly visible in figure 1, the housing prices peaked somewhere in 2008. Because the data is collected throughout the whole year it makes sense to take 2007 as the pretreatment year because the financial crisis hit in September 2008 it might have affected some observations. 2013 is the year when the housing prices were at the lowest in the Netherlands. If wealth would cause short-term changes in health, it should be visible in 2013. So 2013 will be the first post treatment year. In 2017, the housing prices were still not as high as in 2007. Using 2017 as another after treatment year it could show us more about the mid-term effect of wealth on health.

Table two shows the number of observations per year and the size of the treatment and control group. In all the three years the ratio of treatment group to control group is rather consistent and lies around 1:1. There is a trend where the treatment group grows between the years.

Table distribution of the control and treatment groups

|  |  |  |  |
| --- | --- | --- | --- |
| Year | size treatment group | size control group | Total observations |
| 2007 | 835 (46.14%) | 974 (53.84%) | 1809 |
| 2013 | 807 (48.94%) | 842 (51.06%) | 1649 |
| 2017 | 1161 (52.42%) | 1054 (47.58%) | 2215 |

Table three, four and five show the summary statistics of respectively 2007, 2013 and 2017. In all of the three years both the control and treatment group have a fairly high chance of being healthy but the treatment group has a slightly higher chance of being healthy. In the treatment group of 2007, 97% of the people are healthy, in 2013 and 2017 that is 96%. In the control group 95% of the people observed are healthy in 2007 and 2013 and 92% of the people are healthy in 2017. Since the means of the health of both groups are within range of the standard deviations, it can be noted that there is no significant difference between health in the treatment and control group in any of the years. This is important as it indicates that the characteristics across the different treatment groups are successfully balanced.

The average net worth of the treatment group decreases between 2007 and 2013 and then increases between 2013 and 2017. The average net worth of control group increases between 2007 and 2013 and also between 2013 and 2017. Since the house prices also decrease between 2007 and 2013 and increase between 2013 and 2017 this indicates that the average net worth of home owners does behave in line with the housing prices while the net worth of the tenants does not as was expected.

Table summary statistics 2007

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Treatment group | Mean | Standard deviation |
| Health | Yes | 0.97 | 0.18 |
|  | No | 0.95 | 0.22 |
| Net worth | Yes | 275,271 | 329,037 |
|  | No | 25,400 | 75,946 |

Table summary statistics 2013

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Treatment group | Mean | Standard deviation |
| Health | Yes | 0.96 | 0.20 |
|  | No | 0.95 | 0.22 |
| Net worth | Yes | 256,521 | 249,175 |
|  | No | 28,933 | 91,215 |

Table summary statistics 2017

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Treatment group | Mean | Standard deviation |
| Health | Yes | 0.96 | 0.18 |
|  | No | 0.92 | 0.27 |
| Net worth | Yes | 263,053 | 291,670 |
|  | No | 31,530 | 102,177 |

**Methods**

First this paper will use a simple ordinary least squares (OLS) regression. In this OLS regression health will be the dependent variable and wealth will be the independent variable. Health is defined as a categorical variable where someone can be either healthy, with value 1, or unhealthy, with a value of 0. Wealth is the net worth of a person as shown above.

Secondly, this paper uses a difference in difference (DD) analysis to look for a causal effect from wealth to health. Health is defined in the same way as in the OLS regression. A DD analysis is a quasi-experimental design with a treatment group, a control group and a treatment. A DD analysis is used to estimate the effect of a treatment by comparing the differences in the outcomes between before and after the treatment between the treatment and the control group.

The treatment group, control group and the treatment itself will be defined as follows. The treatment this paper uses is the financial crisis of 2008 in which housing prices sharply dropped as is visible in figure 1. It stands to reason that although the sharp drop in housing prices has an effect on the wealth of home-owners, it does not have a direct link to health. It can be assumed that the health of people did not directly cause, or was directly affected by health. The treatment group consists of people who own one or more houses and the control group consists of people who rent a house. Since the participants of the DHS are randomly picked, it can be assumed that the only real difference between the control and treatment group is the ownership of a house. The loss of jobs for example in the financial crisis can be assumed to have the same effect on people who own and people who do not own a house. Therefore, the sharp drop in housing prices caused by the financial can be used as a treatment which only affects the wealth of the treatment group.

The following regression model will be used:

|  |  |  |
| --- | --- | --- |
|  |  | () |

Where Time is the time trend in the control group, Treated is the difference between the two groups pre-intervention and Time \* Diff is the difference in change over time. If the Time \* Diff coefficient is statistically different from zero, there is an effect from the treatment on the dependent variable.

Therefore the hypotheses are:

|  |  |  |
| --- | --- | --- |
|  |  | () |
|  |  | () |

To show that the difference in difference analysis will give the desired causal effect, let be the health of person *i*, in group *g* at period *t* if the person owns a house. Also, let be the health of person *i* in group *g* at period *t* if the person is a tenant. Here the group *g* is either the treatment group (i.e. home-owners) or the control group (i.e. tenants) and the period *i* is either before or after the treatment.

Assume that:

|  |  |  |
| --- | --- | --- |
|  | . | () |

This equation means that in absence of the sudden drop in housing prices caused by the financial crisis, the health of a person is equal to the sum of a time-invariant group effect ( and a time effect that is the same in both the groups (.

Let be a dummy for home-owners and periods. Therefore it is only one when the group is home owners and the period is post-treatment. In the other three cases, is zero. Observed health, , can then be written as

|  |  |  |
| --- | --- | --- |
|  | , | () |

if it is assumed that , a constant. In equation (5), .

Therefore we can get

|  |  |  |
| --- | --- | --- |
|  |  | () |

and

|  |  |  |
| --- | --- | --- |
|  |  | () |

Therefore, the population difference-in-differences is:

|  |  |  |
| --- | --- | --- |
|  | . | () |

Here δ is the causal effect of interest which is estimated through the model described in equation (1). The parameters in the model of equation (1) can be seen in the light of the model described in equation (5) in the following way:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

So is the sum of the time invariant group effect of the tenants and the time effect of the pre-treatment period. therefore it is the average health of the control group in the pre-treatment period.

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

is the time effect of tenants group post treatment minus the time effect of the tenants post treatment. That is equal to the difference in the average health before and after the treatment.

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Therefore is the difference in the time invariant group effect between the home owners and tenants before the treatment. That is the same as to say the difference in average health before the treatment between the two groups.

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

So this is the difference in differences between average health before and after the treatment of the home owners and the tenants.

**Results:**

**Basic OLS results**

A simple OLS regression was conducted first to check whether or not there exists a correlation between health and health in this dataset. With the simple model:

The results of this regression can be seen in table 6. A significant (p < 5%) positive correlation between health and wealth was found as expected. Because health was defined as either healthy of unhealthy the regression tells us something about the chance of being healthy. According to the results a €1000 increase in net wealth corresponds with an increase of 2,81 percentage points to the chance of being healthy. Note that this correlation does not say anything about the possible causal effect.

Table OLS results

|  |  |
| --- | --- |
| Intercept | 0.9476 |
|  | (0.000) |
| Net worth | 2.81 |
|  | (0.016) |
| F-statistic | 5.86 |
|  | (0.0155) |
|  | 0.0010 |

**Difference in Difference Analysis**

As pointed out before in this paper, endogeneity is a problem when considering the effects of wealth on health. Two DD analysis have been done to try to find a causal effect in the short-term and the mid-term of wealth on health. The first analyses the short term effects of wealth on health. It uses the waves from 2007 and 2013. The second one analyses the mid-term effects of wealth on health. For this it uses the 2007 and 2017 waves from the data.

The results for the short term analysis can be seen in table 3, column 2007 - 2013. This analysis does not find a significant causal relationship between health and wealth. This corresponds with the results found by Meer, Miller and Rosen (2003) who also checked for short-term changes in health driven by wealth.

Table DD analysis results

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2007 – 2013 | 2007 -2017 | 2013 - 2017 |
| Intercept | 0.9487 | 0.9487 | 0.9513 |
|  | (0.000) | (0.000) | (0.000) |
| Time | 0.0026 | -0.0265 | -0.0291 |
|  | (0.784) | (0.006) | (0.004) |
| Treated | 0.0190 | 0.0190 | 0.0078 |
|  | (0.049) | (0.063) | (0.470) |
| DID | -0.0111 | 0.0235 | 0.0347 |
|  | (0.422) | (0.088) | (0.015) |
| F-statistic | 1.53 | 9.42 | 7.821 |
|  | (0.203) | (0.000) | (0.000) |
|  | 0.0013 | 0.0070 | 0.006 |

The results from the mid-term analysis can be seen in table 7, column 2007 - 2017. This analysis does find a significant (p<10%) causal relationship between wealth and health using the DD analysis. This suggests that in the Netherlands*,* wealth does have an impact on the chance of being healthy. The coefficient of time is -0.026 which implies that in the time of 2007 until 2017, the chance of being healthy decreased by 2.6 percentage point in the control group. The coefficient of treated of 0.19 means that before the treatment, people who were in the treatment group, the home-owners, had a 1.9 percentage point more chance of being healthy than the people in the control group, the tenants. The interaction coefficient (DID in the table) is 0.023, which implies that after the treatment, people in the treatment group have a 2.3 percentage point more chance on being healthy than their counterparts in the control group. Even though the average housing prices dropped on average between 2007 and 2017, the housing prices were rising again after 2013. This might explain why there is a positive effect instead of the expected negative effect.

To check whether the positive interaction coefficient can be explained by the rise in housing prices between 2013 and 2017 a third DD analysis was conducted between 2013 and 2017. Here the treatment was the rise in housing prices between 2013 and 2017. Therefore the pre-treatment year is 2013 and the post-treatment year is 2017. The results of this analysis can be seen in table 7, column 2013 – 2017. There is a significant (p < 5%) result found which indicates a causal relationship between wealth and health. In this analysis the control group experienced a decrease in the chance of being healthy just as in the analysis of 2007 – 2017. The decrease is 2.9 percentage point between 2013 and 2017. The treated coefficient is not significantly different from zero, so both the treatment and the control group had the same average chance of being healthy. The interaction coefficient implies that after the rise in housing prices, the home owners have a 3.5 percentage point higher chance of being healthy than the tenants.