Debt and Health

A Study of Incapacity Rate and Private Debt in Swedish Municipalities

A thesis submitted in partial fulfillment for the

degree of Bachelor of Science

in the

Department of Economics

School of Business, Economics and Law

University of Gothenburg

Johan Ramne and Benjamin Strömberg

Supervisor: Gustav Kjellsson

Febuary 2017

Abstract

Household debt is rising throughout Sweden, and its health effects are poorly understood. This thesis attempts to establish a link between the average private indebtedness and the incapacity rate—a measure of average paid sick leave—within Swedish municipalities. We present a theoretical causal framework for the effect of debt on health, and review economic literature on the topic. The study finds a small statistically significant negative correlation between average household indebtedness and the incapacity rate, as measured by Försäkringskassan, and which stands opposed to the theory, under the assumption that incapacity rate accurately measures number of days spend sick.

Contents

Introduction

A correlation between the indebtedness of an individual and her health can easily be imagined. Not being able to pay the bills on time could induce serious stress. A lack of funds may be hinder seeking necessary medical treatment, leading to worsening conditions. It is worth noting that a fair amount of research both in the field of health economics and psychology has already been conducted investigating the liaison between indebtedness and poor health. Research has so far indicated there to be a negative correlation between over-indebtedness and psychological health. The link between debt and physical health is not as extensively explored but results up to this date seem to indicate a similar correlation as with mental health. However no causal framework between debt and any of the two health measures has been empirically proven.

This thesis will further the research on the connection between debt and health by examining how average private debt levels aggregated by municipality affects each group's average incapacity rate. First we will review some of this research, along with presenting a brief introduction to the relevant health economic theory surrounding the subject. We will then present the data and regression models used in our study, followed by an analysis of the results and a discussion of their significance.

According to the financial stability report of November 2016, issued by the Swedish Central Bank, the indebtedness of Swedish households has been on the rise since the mid 90's and has recently reached record highs. During the period between July 2010 and July 2016 the mean debt ratio of Swedish households, i.e the ratio between the total debt and total disposable income of a household, has increased from 324 to 343 percent (van Santen and Ölcer, 2016). Although interesting, national financial stability is not a question examined in this thesis. Rather the query at hand is whether indebtedness is a determinant of health measured on a macro level.

Considering the record debt levels and the concerns expressed by the Swedish Central Bank on the issue (Riksbanken, 2016), in combination with the wide spread coverage household debt is receiving

in Swedish media, we believe this inquiry is highly relevant today in order to assess the full societal effect of an indebted population.

The purpose of this study is to examine the link between the indebtedness and health. More precisely this is done by studying the connection between incapacity rate, a measure of the number of sick leave days paid out by the Swedish Försäkringskassan, and private debt in Swedish municipalities during the time period between year 2004 and 2007. A causal analysis is theorized using the Grossman model in health economics and the framework developed by Dackehag et al. (2016). Seeing that private debt has reached all time highs this is a subject which needs further research in order to understand the full societal effect of private debt both on a macro and micro level. We expect this study to fill a gap in the current literature and our results to be explainable within the existing theory on the topic. The two questions this thesis is intended to answer are:

1. *Do average private debt levels in Swedish municipalities effect the health of the population within the region in terms of their incapacity rate?*

2. *Can the observed effects be explained within the framework of the Grossman model and other existing health economic theories?*

The Grossman Model

The framework used to analyze the results in this thesis will be the one of the Grossman model, developed by Michael Grossman (1972). This model is intended to provide answers regarding observed phenomena in health related microeconomics including—but not limited to—the link between an individuals financial strain and health. What will follow is a brief introduction to the Grossman model's ideas and terminology. Only the parts of the model relevant to this thesis are presented.

In a single-period case the utility of the individual is a function of the persons level of health H_t and a composite good Z_t, from now on called the home good, consisting of everything else which would be expected to grant a person utility.

$$U_t = f(H_t, Z_t) \tag{1.1}$$

As in established microeconomic theory the economic person faces a time constraint. In the Grossman model defined as follows.

$$T^{period} = T^W + T^Z + T^H + T^S \tag{1.2}$$

T^{period} defines the total time units available in a given time period, T^W is the time spent working, T^Z is leisure time, T^H is time spend improving health and T^S is time spent sick.

In contrast with regular microeconomic theory inputs to the utility function in the Grossman model cannot solely be obtained by purchasing them on the market. Purchased goods from the market are combined with personal time to obtain utility. The production functions for the inputs to the utility function are defined as follows.

$$H_t = g(H_{t-1}, T_t^H, M_t) \tag{1.3}$$

$$Z_t = h(T_t^Z, J_t) \tag{1.4}$$

Where M_t are market inputs for health, e.g. vaccines and medicine and J_t are market inputs for leisure, e.g. cinema tickets and restaurant visits.

The model also includes a budget constraint:

$$p_M * M_t + p_J * J_t \leq w * T_t^W \tag{1.5}$$

Where w is a person's wage per time unit and p_M and p_J are the market prices of M_t and J_t respectively.

In order to better fit the topic of this thesis the model has to be extended to a multi-period case and to also needs to allow borrowing for future time periods. In the multi-period case the utility function a person seeks to maximize assumes the following form.

$$U = f(H_0, Z_0, H_1, Z_1, ..., H_{\Omega-1}, Z_{\Omega-1}, H_\Omega, Z_\Omega) \tag{1.6}$$

Additionally health is treated as a capital good in the following fashion, β being the rate of depreciation.

$$H_t = H((1 - \beta)H_{t-1}, T_t^H, M_t) \tag{1.7}$$

The budget constraint is modified to include borrowing and saving.

$$p_M * M_t + p_J * J_t + L_t^{repay} \leq w * T_t^W + L_t^{new} \tag{1.8}$$

These new terms L_t^{repay} and L_t^{new} are condensed entries representing the net amount of currency at each time period of existing capital, interest earned, repayment of debts and new loans.

Causal framework - debt and health

Dackehag et al. (2016) has proposed a causal framework explaining the links between financial strain, over-indebtedness and well-being, using in part Bridges and Disney (2010) to affirm their theory. It can be summarized as follows.

In Dackehag et al. (2016)'s model psychological and physical health exhibit a positive correlation between one another. Also the first order auto-correlation of these health measures is considered high. The individual's de-facto financial situation in the current time period is proposed to affect health directly in the next time period as well as the individual's own perceived financial standing in contemporaneous time. An individual's perception of the current situation is also suggested to impact health in the future time period.

Dackehag et al. (2016)'s theory on the topic goes hand-in-hand with the Grossman model. Substantial debt levels could cause stress and inhibit the debtor from choosing the optimal strategy for maximizing his or her health and utility.

To illustrate the Dackehag et al. (2016)-model imagine the following situation. An indebted individual feels stressed over over not being able to repay her loans, say because of being pressured by collectors as discussed by Caplovitz (1974). Nettleton and Burrows (1998) found a positive correlation between mortgage debt levels and psychical health, the fear of losing your home is most likely a powerful stress inducer.

First of all the stress in itself is a serious health condition which increases the risk of a number of diseases. Moreover, the individual, in accordance with the Grossman model, may be forced to sacrifice time used improving health T^H to spend more time working T^W. One could imagine such a trade having a multiplicative effect - T^H being sub-optimal and the additional time at work inducing more stress. On top of this the person might not have sufficient funds to spend on health improving goods M which by definition in the Grossman model would mean poorer health.

1.1   Literature Review

A fairly early study attempting to establish a connection between debt problems and health was conducted by Caplovitz (1974) during the early 1970's in the United States. The study was based on a survey consisting of roughly 1300 interviews with individuals who had been part of debt collection lawsuits due to their indebtedness. About half of the questioned individuals reported that their debt had affected their health negatively and that they had experienced stress regarding the situation.

Another study, this one in a British setting, shows that mortgage indebtedness is affixed with poor mental health health (Nettleton and Burrows, 1998). Findings in their article are based on data from the British Household Panel Survey (BHPS). The BHPS contains information regarding housing, debt, health and several other features through interviews with some 10'000 individuals between the years of 1991 and 1996. Nettleton and Burrows (1998) do not establish any causal relationship but control for physical health, employment status and income.

Similar results have been obtained by Bridges and Disney (2010) also carrying out a study in The UK. Their research paper used data from the so called Families and Child Survey (FACS) between the years of 1999 and 2005. The FACS is a survey designed to extract information on household's health and financial status for families with children and low income. They could conclude that there is a positive correlation between self-reported psychological health and household financial status. However when objective indicators for a households' finances were used to predict psychological health for a the link was much less clear.

Published work up to this date seems to indicate a positive correlation between an individual's economic situation of and his or her mental health. The ultimate measure of this could be argued to be suicide ideation. In addition to the articles presented above, the association between mental health and over-indebtedness is confirmed in this ultimate form in e.g. (Hintikka et al., 1998) and (Meltzer et al., 2010). Both papers relying on survey data, in Finland and the UK respectively.

Furthermore, besides indebtedness exhibiting a strong correlation with psychological health, there also seems to be a link between physical health and over-indebtedness. One study suggesting this is Balmer et al. (2006). This research, based on survey data consisting of interviews with 5611 randomly selected adults on their experiences with, showed that cohorts' debt could be predicted by determining whether they were in a state of long-term illness or disability. The connection between physical health and debt has been confirmed in a Swedish setting, see (Ahlström and Edström, 2015). Ahlström and Edström (2015) showed that the prevalence of health conditions such as stomach ulcers, diabetes, recurrent head aches and high blood pressure to name a few were significantly higher for over-indebted individuals than the benchmark group.

There are a few issues shared by most of the research so far which examines the relationship between debt and health. Firstly, it is difficult to exclude non-observable factors which effect people's health and their risk to to become over indebted, say specific risk-lover behavior or stress-sensitive tendencies. Secondly, research investigating the inverse causal effect i.e. how health influences debt troubles is scarce. Another problem with several studies is that they are based on survey data, hence

making it difficult to draw objective conclusions. Combining this with the fact that this type of research relies on a relatively small sample size propagates the difficulties of inference.

1.2 Contribution

There are a few distinct points which sets this thesis aside from previous research on attempting to explain individuals' health using their indebtedness level.

Regarding the data used in this study we solely use registry data. By not using survey data we mitigate possible biases connected with that data form. Also, it allows us to have a large sample size. Furthermore the way we define indebtedness differs from existing literature. Papers to this date generally examine the association between some form of over-indebtedness, for which there is no universal definition, and health. To contrast this we study average debt levels, defined as the ratio between private assets and debt within a region. Defining debt this way gives an objective definition of indebtedness and captures a broader spectrum of indebted individuals which may or may not be affected in the same way as over-indebted people. This hints at the last point, by studying average indebtedness in regions we use aggregate form data. Obviously aggregate data has some disadvantages, however it may allow us to see patterns on a macro level complementing the current published research.

Hypothesis

The hypothesis of the study in this thesis is that incapacity rate is a good proxy variable for T^S in the Grossman model. If this hypothesis is true we also expect to observe a positive correlation between incapacity rate and private debt. A result which would be in line with the current literature base, expanding empirical evidence to a macro level and encompassing all debt levels.

Method

2.1 Data

2.1.1 Incapacity Rate

We have chosen to use the variable incapacity rate as our proxy measure for time spent sick among Swedes. The incapacity rate is the ratio of the number of sick days compensated by Försäkringskassan, Sweden's Social Security Authority, to the number of insured persons within the given region. For most working Swedes their company covers the first 14 days of sick leave, and it is first beyond this point that Försäkringskassan steps in (Försäkringskassan, 2016b). The data does therefore not generally capture short-term sickness among employees. Unemployed persons registered with the Swedish Public Employment Service can apply for sick pay if their sickness makes them unable to apply for jobs, or unable to accept one (Försäkringskassan, 2016c).

Employees are required to send in a medical certificate before receiving money from Försäkringskassan; unemployed people are required to do so within 7 days of applying for sick pay. The incapacity rate also includes sick days due to a work related injury, as well as compensation to persons younger than 30 who cannot join the labor market for at least one year due to some disability or illness (Försäkringskassan, 2016a). The data is compiled by Försäkringskassan themselves.

The incapacity rate is one of the most granular objective measures of health that is publicly accessible in Sweden. Though it is an imperfect measure, as health has different meanings in different contexts, we believe it is the closest obtainable measure of T^S in the Grossman model. The incapacity rate does not say much about the particular illnesses and circumstances that are prevalent within different municipalities, but this can be circumvented through clever choice of regression model, as we will show later.

2.1.2 Indebtedness

Our independent variable of interest is average indebtedness in a region, which we have defined as the ratio of all assets, real and financial, to all owed debts. The data-set includes only individuals registered as owning any assets or owing debts, thus excluding small children and unregistered migrants. The data was collected by the Swedish Tax Authority and compiled by Statistics Sweden for the years 2004–2007. The underlying reason for the Swedish Tax Authority registering the assets and debts of individual citizens was for the collection of a wealth tax, however this tax was abolished in January 2007, and the registry was discontinued (Regeringen, 2007).

The indebtedness data constricts the study's granularity and range, as the publicly available data is on a municipal level and can at best be obtained for the age intervals 20–29, 30–49 and 50–64, from 2004 to 2007, along with being differentiable by gender. The assets are calculated at market prices on December 31 for the respective years. Assets such as housing, stocks, bonds and bank savings are included in the data, however other assets such as cars, boats and artwork are not. We argue that the data-set is inclusive enough, and that the excluded asset classes have a negligible effect on the result as a whole.

2.1.3 Data Cells

We have chosen to divide Sweden into cells consisting of municipality, gender, and three age intervals: 20–29, 30–49, and 50–64. These choices are mainly due to the limitations in the indebtedness data. To fit the incapacity rates to these cells, we have combined values weighted by the proportion of the population within their assigned age interval. Thus, the incapacity rate of 30–49 interval for each municipality and gender is calculated as:

$$\text{incap}_{30-49,m,t,g} = \text{incap}_{30-39,m,t,g} * \theta_{30-49,m,t,g} + \text{incap}_{40-49,m,t,g} * \left(1 - \theta_{30-49,i,t,g}\right) \tag{2.1}$$

where incap is the incapacity rate, m is counter for the municipalities, t denotes the year from 2004–2007, g is gender dummy that equals one for women and zero for men, and

$$\theta_{30-49,m,t,g} = \frac{\text{pop}_{30-39,m,t,g}}{\text{pop}_{30-39,m,t,g} + \text{pop}_{40-49,m,t,g}} \tag{2.2}$$

where $\text{pop}_{30-39,i,t,g}$ and $\text{pop}_{40-49,i,t,g}$ are the number of persons in the municipality within the given age-interval, year and gender.

The incapacity rate for the age-interval 50-64 is calculated in the same way, using $\mathsf{incap}_{50-59,i,t,g}$ and $\mathsf{incap}_{60-64,i,t,g}$ along with the respective population variables.

2.1.4 Control Variables

In the model we have included the regional unemployment rate, population density and a measure of education as time-varying control variables. The unemployment data is provided by the Swedish Public Employment Service, while the population density—measured as citizens per square kilometer—along with the education data are compiled by Statistics Sweden. Our education variable is the percentage of persons in the region with post-gymnasial education.

We know that unemployment has a direct effect on the incapacity rate, as it means Försäkringskassan also pay for the first 14 days of sickness, where the employer would otherwise have been paying. Further, unemployment on an individual level is associated with worsened mental health (Murphy and Athanasou, 1999; Kieselbach et al., 2006). An increase in population density may signal an increased pressure on health care services, as well as increasing the probability of an infectious disease spreading.

Education has been shown to have effects on a broad number of health measures (Bloom, 2005; Fischer et al., 2013), as well as being a determinant of what type of work is generally available to the population. A higher number of years in education—especially beyond the gymnasial level—generally leads to less physically demanding work, which may decrease incapacity rate.

2.2 Model Choice

With our choice of variables, the linear model to be estimated is:

$$
\begin{aligned}
\mathsf{incap}_{age,m,t,g} = {} & \beta_0 + \beta_1 \mathsf{debt}_{age,m,t,g} + \beta_2 \mathsf{unemp}_{m,t} + \beta_3 \mathsf{educ}_{age,m,t,g} + \beta_4 \mathsf{den}_{m,t} + \beta_5 \mathsf{YEAR2004} \\
& + \beta_6 \mathsf{YEAR2006} + \beta_7 \mathsf{YEAR2007} + \beta_8 \mathsf{Woman} + \beta_9 \mathsf{AGE2029} + \beta_{10} \mathsf{AGE5064} + U_{age,m,t,g}
\end{aligned} \tag{2.3}
$$

where incap is the the incapacity rate; debt is the indebtedness variable; unemp denotes the unemployment; educ is our education variable; den is the population density; and YEAR2004, YEAR2006, YEAR2007, Woman, AGE2029 and AGE5064 are dummy variables. We have chosen men in the age-interval 30-49 as the benchmark group, with 2005 as the benchmark year. The subscripts denote the age interval (age), municipal number (m), time (t) and gender (g) for the observation. These subscripts will henceforth be excluded for easy of reading in places where they do not add clarity.

Which estimation model gives the best results depends on the attributes of the data. This section describes our tests for different characteristics and problems within the data that affect the efficiency and validity of inference statistics in different estimation models.

2.2.1 High Persistence

Nearly all the time series data used in this study are intuitively at risk of being highly persistent over time, i.e having a strong dependence between x_t and x_{t-1}. This high persistence is problematic as it voids the law of large numbers and the central limit theorem—which require independent trials and variables respectively—that controls some of the errors that can occur in the model due to violations of the Classic Linear Model (CLM) assumptions (Wooldridge, 2014).

To look for high persistence, we tested the correlation between each variable and their time-lagged counterparts. The results are found below in table 2.1. By convention $|\rho| > 0.9$ is judged to indicate high persistence. We can clearly see that all variables are highly persistent.

Variable	Autocorrelation
Incapacity Rate	0.9976
Indebtedness	0.9795
Unemployment	0.9089
Education	0.9948
Population Density	1.0000

Table 2.1: The ρ of AR(1) for the model variables

2.2.2 Multicollinearity

Next, we investigate if there is high multicollinearity between the variables: that is, if the variance of any one of our explanatory model variables can to a high degree be explained by the variance in the others. High multicollinearity for a variable leads to larger variance in its estimated coefficient, decreasing the likelihood of a statistically significant result.

We tested the multicollinearity using a *variance inflation factor* (VIF) test, the results of which are seen in table 2.2. As a crude rule of thumb, you can assume that multicollinearity will be a problem for coefficients with **VIF** > 10. However, this is only a part of the picture, as the variance of a coefficient β also depends on the variance in the dependent variable, as well as the variance in the explanatory variable itself. As none of the model variables have terribly high VIF's, we will not exclude any of them from the regressions.

Variable	VIF	1/VIF
debt	6.59	0.1518
unemp	1.40	0.7156
educ	1.70	0.5878
den	1.29	0.7757
YEAR2004	1.51	0.6610
YEAR2006	1.56	0.6393
YEAR2007	1.79	0.5601
Woman	1.43	0.7013
AGE2029	3.08	0.3243
AGE5064	3.32	0.3011

Table 2.2: Variance inflation factors for the model variables

2.2.3 Serially Correlated Errors

In addition to testing the model variables for high persistence and multicollinearity, we tested our stated model (2.3) for first order serially correlated errors. This was done by calculating the residuals U_t of a OLS regression, and then running the regression:

$$U_t = \beta_0 + \rho U_{t-1} + \beta_2 \mathsf{debt}_t + \beta_3 \mathsf{unemp}_t + \beta_4 \mathsf{educ}_t + \beta_5 \mathsf{den}_t + \beta_6 \mathsf{YEAR2006}$$
$$+ \beta_7 \mathsf{YEAR2007} + \beta_8 \mathsf{Woman} + \beta_9 \mathsf{AGE2029} + \beta_{10} \mathsf{AGE5064} + e_t \tag{2.4}$$

We suspected heteroskedasticity in this regression and chose to calculate standard errors that are robust towards this. The results of are reported in table 2.3 on the next page, with the main point of interest being $\hat{\rho} = 0.9286$, indicating that we do have serially correlated errors. Thus tests for heteroskedasticity will be applicable only after we have dealt with the serial correlation (Wooldridge, 2014).

Their are several ways to combat the issues caused by high persistence and serially correlated errors, each with its own benefits and drawbacks. We will look at two such solutions: first-differenced transformation and fixed effects transformation with clusters.

Variable	Coefficient	(Std. Err.)	P > |t|
U_{t-1}	0.9286	(0.0046)	0.000
debt	1.5591	(0.4638)	0.001
unemp	-31.3516	(4.2776)	0.000
educ	-0.9414	(0.3748)	0.012
den	0.0003	(0.0001)	0.000
YEAR2006	-0.1223	(0.0906)	0.177
YEAR2007	-0.3212	(0.0973)	0.001
Woman	-0.1021	(0.0854)	0.232
AGE2029	0.0665	(0.2561)	0.795
AGE5064	-0.4289	(0.1554)	0.006
Intercept	0.1456	(0.1718)	0.397

Dependent Variable: U_t

Table 2.3: Results from test for serial correlation

2.2.4 First-differenced Transformation

The first-differenced transformation regresses the change in the dependent variable on the change in the explanatory variables between time periods. If we have $\rho = 1$ for a variable x_t, taking the first-difference $x_t - x_{t-1} = \Delta x_t$ generates a new weakly dependent variable instead, assuming x_t follows an AR(1) model.

It makes intuitive sense to use a first-differenced transformation if the error $U_{age,m,t,g}$ can be broken up into a composite error $U_t = a + V_t$, where a is a constant specific to the cell that is invariant in time. This unobserved effect can for example be due to the type of work available in a region, or how specific types of education (lawyer, welder, researcher etc.) are more prevalent in some areas than others. The real usefulness of a first-differenced transformation is that it removes a from the equation as $\Delta a = 0$. Our full first-differenced model is then:

$$\Delta \text{incap} = \beta_0 + \beta_1 \Delta \text{debt} + \beta_2 \Delta \text{unemp} + \beta_3 \Delta \text{educ} + \beta_4 \Delta \text{den} + \beta_5 \text{YEAR2006} + \beta_6 \text{YEAR2007} + \Delta V$$

$$(2.5)$$

Note that the dummy variable YEAR2004 has to be dropped to in response to the differencing. The first-differenced tranformation has the drawback of not being able to calculate any effects that are constant over time—these disappear in the same way as a does. We can therefore not directly observe how gender and age-group affect the incapacity rate with this model; however, partitioning the municipalities along age and gender uses the data to a fuller extent by giving us greater granularity and more observations.

If the original time series contained serial correlation of type AR(2), then first-differencing will not have completely removed the problem. We test ΔV for serial correlation with the results, calculated with standard errors robust to heteroskedasticity, shown in table 2.4. We see that the coefficient for ΔV_{t-1}, $\rho = 0.2286$, is the only statistically significant explanatory variable in the regression, and we can quite safely make the assumption that we have removed most of the serial correlation in the errors as $\rho << 1$. Under this assumption the law of large numbers and the central limit theorem allow us to use t-statistics and F-tests in large samples. When obtaining ΔV_{t-1} we lose yet another year of observations, and have to exclude the dummy YEAR2006.

Variable	Coefficient	(Std. Err.)	P > |t|
ΔV_{t-1}	0.2286	(0.0257)	0.000
Δdebt	-0.2356	(1.1506)	0.838
Δunemp	3.8914	(14.3823)	0.787
Δeduc	-0.2432	(7.1721)	0.973
Δden	0.00129	(0.0031)	0.675
YEAR2007	-0.0111	(0.0878)	0.900
Intercept	0.0170	(0.1006)	0.866

Dependent variable: ΔV_t

Table 2.4: Test for serial correlation in ΔV_t

2.2.5 Fixed Effects Transformation

The fixed effects transformation is a different method for removing the effects of high persistence and removing unobserved constant cell-specific effects. It does this by regressing the difference between each sample point and the sample mean for all the variables that are highly persistent. That means instead of using a highly persistent variable x, we use $x_{m,t} - \bar{x}_m$, where $\bar{x}_m = T^{-1} \sum_{t=1}^{T} x_{m,t}$. The full equation then becomes:

$$
\begin{aligned}
\text{incap} - \overline{\text{incap}} =& \beta_0 + \beta_1(\text{debt} - \overline{\text{debt}}) + \beta_2(\text{unemp} - \overline{\text{unemp}}) + \beta_3(\text{educ} - \overline{\text{educ}}) + \beta_4(\text{den} - \overline{\text{den}}_m) \\
& + \beta_5\text{YEAR2004} + \beta_6\text{YEAR2006} + \beta_7\text{YEAR2007} + a - \bar{a} + V - \overline{V}
\end{aligned}
\tag{2.6}
$$

As in the first-differenced model we have a composite error term $U_t = a + V_t$, and as a is constant, we have $\bar{a} = a$ which cancels out the term. Using the notation $\ddot{y}_{m,t} = y_{m,t} - \bar{y}_m$ the model condense to:

$$\ddot{\text{incap}} = \beta_0 + \beta_1\ddot{\text{debt}} + \beta_2\ddot{\text{unemp}} + \beta_3\ddot{\text{educ}} + \beta_4\ddot{\text{den}} + \beta_5\text{YEAR2004} + \beta_6\text{YEAR2006} + \beta_7\text{YEAR2007} + \ddot{V}$$

(2.7)

The fixed effects model does not on its own negate the effects of serial correlation in the error term as effectively as first-differencing (Wooldridge, 2014), making the inference tests much more sensitive to its presence. However, when the error terms are serially uncorrelated fixed effects model is more efficient. One problem that both fixed effects and first-differencing have is that they require large variation in the key explanatory variables across time: if we have no variation in the debt level between the given years, then the correlation we are interested in examining will be subsumed to a and cancels out.

Their is no good way of reverse engineering the error term V from $\ddot{V}$. Instead, we have to rely on the test for serial correlation in the linear model. As we made the deduction that the linear model contained serial correlation, the same will be true for the fixed effects model. In this case, the coefficient estimates will according to Wooldridge (2014) still be consistent, though possibly not unbiased. However, the variance estimation for the coefficients will be underestimated when the serial correlation is positive, meaning we are more likely to make a type I error that the confidence level suggests.

Fixed Effects with Clustered Standard Errors

One way of dealing with serial correlation and heteroskedasticity in a fixed effects regression is to cluster the standard errors. By weakening the assumptions governing the regression, specifically the independence of the intra-year error terms, and assuming dependence within groups and independence between them, we can obtain more accurate standard errors (Institute for Digital Research And Education, 2016). In this study we have chosen to estimate FE-regressions with the standard errors cluster at two different levels: according to municipality, and according to cell.

18

Results

In this chapter the results of the empirical work of the thesis are presented. Regression outputs are shown in tables which in turn are explained thoroughly.

3.1 Regression Analysis

3.1.1 Linear Model

The first run regression explores the connection between the incapacity rate and indebtedness using the multivariate linear model previously specified in section 2.2.2. All variables inserted into the model are non-transformed resulting in the following equation.

$$
\begin{aligned}
\text{incap} = {} & \beta_0 + \beta_1\text{debt} + \beta_2\text{unemp} + \beta_3\text{educ} + \beta_4\text{den} + \beta_5\text{YEAR2004} + \beta_6\text{YEAR2006} \\
& + \beta_7\text{YEAR2007} + \beta_8\text{Woman} + \beta_9\text{AGE2029} + \beta_1 0\text{AGE5064} + U
\end{aligned} \tag{3.1}
$$

As can be deduced from table 3.1, on the next page, all variables in the untransformed linear model are statistically significant at the five- and one-percent significance levels including the variable of interest debt, revealing a t-statistic of -11.76. The coefficient β_1 should be interpreted as the following. A municipality which increases its mean debt-ratio by one percentage point in absolute terms is expected to experience a on average decrease in incapacity rate by 0.132055 days amongst its population. Also it can be noted that the R-squared in this specification is close to one, 0.9115 meaning that the independent variables account for ninety-one percent of the variation in incapacity rate in the population.

Dependent variable: incap

Variable	Coefficient	(Std. Err.)	t-statistic	P > |t|	95% Confidence Interval	
debt	-13.2055	(1.1225)	-11.76	0.000	-15.4059	-11.0051
unemp	341.0139	(15.3221)	22.26	0.000	310.9779	371.0499
educ	-61.3490	(1.6892)	-36.32	0.000	-64.6603	-58.0378
den	0.0007	(0.0002)	2.94	0.003	0.0002	0.0012
YEAR2004	0.7325	(0.3501)	2.09	0.036	0.0462	1.4189
YEAR2006	0.5766	(0.3427)	1.68	0.092	-0.0952	1.2484
YEAR2007	1.7330	(0.3688)	4.70	0.000	1.0100	2.4561
Woman	24.8078	(0.3227)	76.88	0.000	24.1752	25.4403
AGE2029	-20.6442	(0.4010)	-51.48	0.000	-21.4302	-19.8581
AGE5064	43.2405	(0.4330)	99.87	0.000	42.3917	44.0892
Intercept	35.4818	(0.8207)	43.23	0.000	33.8728	37.0908

Number of Observations: 6960

$F(10, 6949) = 5143.37$

Prob > F = 0.000

R-squared: 0.9115

Table 3.1: Results of linear regression

At a first glance these results seems statistically compelling. However, considering that the dependent variable and all independent variables, aside from the dummy variables, seem to be highly persistent along with the fact that residuals exhibit serial correlation according to the tests in section 2.2.3. These results are disregarded and will not be used for further inference analysis.

3.1.2 First-differenced Model

In our first attempt to combat the problems of high persistency and serial correlations in the error term we regressed the data using a first-difference model. As explained in section 2.2.4 this is done by a first-difference transformation of all the highly persistent variables: incapacity rate, debt, unemployment, education and population density. The age and gender dummy variables are dropped in this model as effects constant over time cancel out. The equation being regressed is the following:

$$\Delta\text{incap} = \beta_0 + \beta_1\Delta\text{debt} + \beta_2\Delta\text{unemp} + \beta_3\Delta\text{educ} + \beta_4\Delta\text{den} + \beta_5 2006 + \beta_6\text{YEAR2007} + \Delta V$$

$$(3.2)$$

Dependent variable: Δincap

Variable	Coefficient	(Std. Err.)	t-statistic	P > |t|	95% Confidence Interval	
Δdebt	-1.9419	(0.8688)	-2.24	0.025	-3.6451	-0.2388
Δunemp	24.8092	(10.8679)	2.28	0.022	3.5036	46.1148
Δeduc	-16.1827	(5.4030)	-3.00	0.003	-26.7747	-5.5906
Δden	-0.0007	(0.0029)	-0.24	0.813	0.0064	0.0050
YEAR2006	-0.3615	(0.0989)	-3.66	0.000	-0.5553	-0.1676
YEAR2007	-0.5134	(0.1115)	-4.60	0.000	-0.7320	-0.2948
Intercept	-0.5945	(0.0653)	-9.10	0.000	-0.7226	-0.4664

Number of Observations: 5220

$F(6, 5213) = 18.45$

Prob > F = 0.000

R-squared: 0.0233

Table 3.2: Results of first-differenced regression

The first-differenced regression outputs can be observed in table 3.2. All independent variables are statistically significant at the five percent significance level except Δden. However neither our variable of interest Δdebt, which reveals a t-statistic of -2.24, nor Δunemp are significant at the one percent level.

Coefficients for first-differenced variables are interpreted in the same way as for the regular linear regression. Subtracting the value from the previous time period does not effect the the β-values of the equations more than in a statistically technical manner. This means that for β_1, assuming all else equal, an increase in the average debt ratio for individuals in a given municipality by one percentage point in absolute terms is expected to decrease the average incapacity rate in that municipality by 0.019419 days. Even though β_1 is statistically significant it can be argued to be practically insignificant considering that the average incapacity rate across all groups during the time interval is about 46 days.

3.1.3 Fixed Effects with Clusters

To ensure robustness of the findings of the first difference regression two regressions using fixed effects transformation with clusters have been run. As a reminder from section 2.2.5 the regressed equation is the following.

$$\ddot{\text{incap}} = \beta_0 + \beta_1\ddot{\text{debt}} + \beta_2\ddot{\text{unemp}} + \beta_3\ddot{\text{educ}} + \beta_4\ddot{\text{den}} + \beta_5\text{YEAR2004} + \beta_6\text{YEAR2006} + \beta_7\text{YEAR2007} + \ddot{V}$$

(3.3)

Table 3.3, on the next page, shows the outputs of the first regression which is clustered with respect to each cell. That is age, gender and municipality of residence. The next table, table 3.4, shows the results of the second clustered regression, where each cluster represents a municipality.

In both of the fixed effects regressions the variable of interest $\ddot{\text{debt}}$ is significant at the five percent level. The sign and magnitude of the coefficient β_1 remains the same as in the first difference regression strengthening our findings. Additionally it is still relatively small further indicating that the effect of average private debt in Swedish municipalities on the populations incapacity rate is small. What should be noticed is that the variable which controls for education is not significant even on a ten percent level in any of the fixed effect regressions. Otherwise the remaining controls exhibit similar characteristics as in the first difference model.

We would like to point out that the very low R-squared values of the all the regressions with transformed variables is not of concern. The purpose of the thesis is to investigate the link between indebtedness and incapacity rate, not to fully explain it.

Dependent variable: in̈cap

Variable	Coefficient	(Std. Err.)	t-statistic	P > |t|	95% Confidence Interval	
debt	-3.4374	(1.4014)	-2.45	0.014	-6.1860	-0.6888
unëmp	17.913	(16.8530)	1.06	0.288	-15.1413	50.9673
educ	-33.1377	(6.0646)	-5.46	0.000	-45.0325	-21.2430
den	0.0013	(0.0040)	0.33	0.743	-0.0065	0.0091
YEAR2004	0.5454	(0.0677)	-8.05	0.000	0.4125	0.6782
YEAR2006	-1.0031	(0.0.1122)	-8.94	0.000	-1.2231	-0.7831
YEAR2007	-2.0982	(0.2218)	-9.46	0.000	-2.5332	-1.6632
Intercept	55.7912	(1.7389)	32.08	0.000	52.3807	59.2018

Number of Observations: 6960
Number of Clusters: 1740
$F(7, 1739) = 124.37$
Prob > F = 0.000
R-squared within: 0.2516
R-squared between: 0.0446
R-squared overall: 0.0441

Table 3.3: Results of fixed effects regression with cell clusters

Dependent variable: in̈cap

Variable	Coefficient	(Std. Err.)	t-statistic	P > |t|	95% Confidence Interval	
debt	-3.4374	(1.4701)	-2.34	0.020	-6.3308	-0.5440
unëmp	17.913	(24.6463)	0.73	0.468	-30.5959	66.4219
educ	-33.1377	(6.5687)	-5.04	0.000	-46.0662	-20.2093
den	0.0013	(0.0032)	0.41	0.684	-0.0050	0.0076
YEAR2004	0.5454	(0.0838)	6.51	0.000	0.3804	0.7103
YEAR2006	-1.0031	(0.1539)	-6.52	0.000	-1.3060	-0.7002
YEAR2007	-2.0982	(0.3156)	-6.65	0.000	-2.719422	-1.4770
Intercept	55.7913	(1.9923)	28.00	0.000	51.8699	59.7127

Number of Observations: 6960
Number of Clusters: 290
$F(7, 289) = 79.66$
Prob > F = 0.000
R-squared within: 0.2516
R-squared between: 0.0446
R-squared overall: 0.0441

Table 3.4: Results of fixed effects regression with municipality clusters

3.1.4 Analysis

The results from the first-differenced regression and fixed effects regressions confirm what the CLM initially indicated, but contradict our expectations of this study. Earlier we hypothesised that incapacity rate was a satisfactory proxy variable for health or time spend sick, T^S in the Grossman model. Also, the review of current literature showed that health, especially psychological well-being to be positively correlated with high debt levels. In light of our original hypothesis on incapacity rate, together with what the literature base seemed to indicate, we believed that we were going to observe the opposite relationship between average private debt and incapacity rate from what the regressions showed.

Now, we must ask the question: Have we contradicted research to this date or have we observed a different relationship all together?

Published research agrees on that over-indebtedness is linked to decaying mental health, foremost in the form of high stress levels. Considering this we find it hard to believe that we have, even though small, showed the opposite correlation in this study. Using the Grossman model we would like to propose two different interpretations which could be responsible for the observed results either individually or in tandem.

Incapacity Rate and Time Spent Sick

A first possible explanation is that one of our initial premises of the thesis is wrong. Incapacity rate may not be a relevant proxy for T^S in the Grossman model. Instead it may be combination of several components in the time constraint, T^S, T^Z and T^H: sick time, leisure time and time spent improving health. Dackehag et al. (2016) theorized that an indebted person has to sacrifice time spent on leisure and improving health for time spent working in order to pay previously taken loans. Then, an individual may be forced to work when he or she would need to spend time improving health resulting in deteriorating well being according to the Grossman model (Grossman, 1972). If this is the case a decrease in incapacity rate would mean a worsened health.

Healthy Debt and Interactions

When motivating the study in this thesis we pointed out that current literature examines over-indebtedness and its link to health. As the variable of interest in our study is average debt measured regionally we would like to take a step back and briefly examine what this assessment of debt actually is. Debt at reasonable levels, whatever that may be, does not necessarily have to lead down the destructive path presented by Dackehag et al. (2016). Most often an increased private debt is associated with the

purchase of a home, which definitely could be considered a health improving action and investment. Hence the negative correlation showed in other studies may only apply above a certain threshold or in combination with other factors specific for individuals with higher amounts of debt.

After these short causal speculations we would like to remind the reader that as we have not identified any exogenous variation in the variable of interest. This being the case our speculations about causality remain just that.

3.2   Limitations

The main limitation of this study is that we have not identified any exogenous change in our variable of interest, the indebtedness. Without this, we cannot construct the study as a natural experiment with a control group and treatment group, and we cannot make any inference about causality. Instead, we aim to add to the literature on the subject by trying to establish a correlation between indebtedness and the incapacity rate.

Another limitation is the small time span we are investigating—especially with regards to the first-differencing and fixed effects regressions. Both these methods require large variance over time in both the dependent and the independent variables to be able to establish any reliable coefficients, as constants get canceled out. Over a four-year period the variance of both the incapacity rate and the indebtedness ratio may be to small. The small time span also gives the results less external validity going forward, as we cannot rule out that the results are due to something specific to the mid-2000s, instead of being something intrinsic in society.

We have to be very careful when making inferences about individuals when the data used is in aggregate form, so as to not fall into the ecological fallacy, where a correlation disappears when data on an individual level is examined (Robinson, 1950). This is especially true as when have no knowledge of either the distributions of the indebtedness or the incapacity rate within a municipality. It is entirely possible that how assets and debt are distributed within the community affects the incapacity rate.

Discussion

4.1 Conclusions

Due to the issue of serial correlation in the error term, we believe that the first-differenced model gives the most accurate results. The coefficient for debt is statistically significant at an five percent significance level; however, the effect is negative—contrary to our hypothesis—and very small. The results in the first-differenced regression are supported by those in the fixed effects regressions. All regressions establish a negative correlation between the average incapacity rate and the average indebtedness level with a municipality.

Due to the data being in aggregate form we have to be very cautious when making inference about effects on an individual level. As the we lack exogenous variance in the indebtedness variable, we also have to refrain from making any empirical deductions about the direction of causality. We can however say that if the incapacity rate is a good approximation of time sick in the Grossman model, then our results stand opposed to the theoretical framework.

We see two reasonable explanations for the disparity between our results and the Grossman model: we believe that either our initial assumption that the incapacity rate approximates T^S (time spent sick) was false, and that it may include time spent improving health and leisure times as well; or the interaction between incapacity rate and health is concave, implying the existence of "healthy debt". Also there non-observable factors could be at play while determining how specific groups are effected by various amounts of debt.

The external validity of our result is further hampered by the fact that the data is collected over a relatively short time-frame. However, the size of our sample gives our tests the power to detect relatively small effects. Overall, we believe that further studies with greater time-frames need to be conducted before any conclusions on the correlation between indebtedness and health are made.

4.2 Further Studies

We would like to propose a few areas in which further work on the subject would be of benefit. The Swedish Central Bank is currently in possession of a anonymized panel data-set over the indebtedness of all Swedish citizens, constructed together with the credit-information company UC (Winstrand and Ölcer, 2014; van Santen and Ölcer, 2016). The data contains values for at least 2014 and 2016, and the individuals can be combined into municipal averages. This data could be used to further test the link between incapacity rate and indebtedness in a different time span, and would not have to be as restricted to the same age-intervals as our study has been. Alternatively, a similar data-set could feasible be constructed together with a credit-information company.

The second area of interest is attempting to confirm the correlation between health and debt on an individual level by collecting some type of objective health measure from a large sample of individuals, and then comparing this to the indebtedness-level, gathered through the services of a credit-information bureau. This kind of study would have the additional benefit of isolating the effect of unemployment, but would not be able to account for education if using a fixed effects or first-differenced transformation.

Bibliography

Ahlström, R. and Edström, S. (2015). Överskuldsättning och ohälsa. Technical report, Konsumentver-
   ket.

Balmer, N., Pleasance, P., Buck, A., and Walker, H. C. (2006). Worried sick: The experience of
   debt problems and their relationships with health, illness and disability. *Social Policy and Society*,
   5:39–51.

Bloom, D. E. (2005). Education, health, and development. *Comparative Education Review*, 49(4).

Bridges, S. and Disney, R. (2010). Debt and depression. *Journal of Health Economics*, 29:388–403.

Caplovitz, D. (1974). *Consumers In Trouble: A Study Of Debtors In Default*. Free Press, New York,
   USA.

Dackehag, M., Ellegård, L. M., Gerdtham, U., and Nilsson, T. (2016). Ekonomisk knapphet, skuldsät-
   tning och hälsa. Technical report, Kronofogden.

Fischer, M., Karlsson, M., and Nilsson, T. (2013). Effects of compulsory schooling on mortality:
   Evidence from sweden. *International Journal of Environmental Research and Public Health*, 10.

Försäkringskassan (2016a). Activity compensation in the event of disability. `https:`
   `//www.forsakringskassan.se/privatpers/funktionsnedsattning/`
   `aktivitetsersattning_och_sjukersattning_lank/aktivitetsersattning`
   [Accessed: 2016-12-20].

Försäkringskassan (2016b). Sickness benefit for employees. `https://www.`
   `forsakringskassan.se/privatpers/sjuk/anstalld/sjukpenning` [Accessed:
   2016-12-20].

Försäkringskassan (2016c). Sickness benefit for job seekers. `https://www.forsakringskassan.se/privatpers/sjuk/arbetssokande/sjukpenning` [Accessed: 2016-12-20].

Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy*, 80(2):223–255.

Hintikka, J., Kontula, O., Saarinen, P., Tanskanen, A., Koskela, K., and Viinamäki, H. (1998). Debt and suicidal behaviour in the finnish general population. *Acta Psychiatrica Scandinavica*, 98:493–496.

Institute for Digital Research And Education (2016). Regressions with stata: Chapter 4—beyond ols. `http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter4/statareg4.htm` [Accessed: 2016-12-22].

Kieselbach, T., Winefield, A. H., Boyd, C., and Andersson, S. (2006). *Unemployment and Health: International and Interdisciplinary Perspectives*. Australian Academic Press, Brisebane, Australia.

Meltzer, H., Bebbington, P., Brugha, T., Jenkins, R., McManus, S., and Dennis, M. S. (2010). Personal debt and suicidal ideation. *Psychological Medicine*, 41:771–778.

Murphy, G. C. and Athanasou, J. A. (1999). The effect of unemployment on health. *Journal of Occupational and Organizational Psychology*, 72:83–99.

Nettleton, S. and Burrows, R. (1998). Mortgage debt, insecure home ownership and health: an exploratory analysis. *Sociology of Health Illness*, 20(5):731–753.

Regeringen (2007). Regeringens proposition 2007/08:26 slopad förmögenhetsskatt m.m. Technical report, Regeringen.

Riksbanken (2016). Banker och hushåll alltmer sårbara. Press release. `http://www.riksbank.se/sv/Press-och-publicerat/Pressmeddelanden/2016/Banker-och-hushall-alltmer-sarbara/` [Accessed: 2016-12-20].

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 30(3):351–257.

van Santen, P. and Ölcer, D. (2016). Svenska hushållens skuldsättning – uppdatering för 2016. Technical report, Riksbanken.

Winstrand, J. and Ölcer, D. (2014). Hur skuldsatta är det svenska hushållen? Technical report, Riksbanken.

Wooldridge, J. M. (2014). *Introduction to Econometrics*. Cengage Learning, Hampshire, UK.