Analysing the contribution of health to the gender wage gap in Germany using a Quantile regression and Oaxaca-Blinder decomposition technique

Research report (11.729 words)

Sanne van Deudekom
Student number: 432158
Supervisor: Teresa Marreiros Bago d'Uva
Rotterdam, 29-6-2016

ABSTRACT

Analysing the contribution of health to the gender wage gap in Germany using a Quantile Regression and Oaxaca-Blinder decomposition technique

The gender wage gap is a topic that has been discussed and studied extensively. Only a few studies have included health factors as explanatory variables for the wage gap, whilst it has been shown that health does affect income. This study uses the GSOEP dataset to research the contribution of health to the wage gap in Germany using 20 waves of the period 1993-2013. Actual health impairment indicators and self-assessed health values are used as health variables and they are amongst others, regressed on the logarithm of hourly wage. Quantile regression is first applied to investigate the contribution of health to wages separately for males and females. Subsequently, the Oaxaca-Blinder decomposition technique is applied for decomposing the wage differential of this sample into detailed information of the contribution of each explanatory variable individually. This is subdivided into the explained differential, the wage gap due to differences in characteristic distributions between genders, and an unexplained differential. The latter constitutes differences in returns to characteristics, and is named as the actual wage gap. This is the gap which might expose discrimination. The results show that the females in this sample on average have a lower self-assessed health than males. For both genders self-assessed health positively contributes to wages, where this is stronger for males than females. This effect is found being the strongest in the lower income levels. The health impairment variables negatively contribute to wages, but only significantly in the higher income levels. Decomposition of the wage differential shows that the health impairment variables have no contribution to the unexplained wage gap. Self-assessed health does have a significant contribution to this gap, however it only entails 6.4% of the total unexplained gap.

Keywords: decomposition, gender wage gap, health, Oaxaca-Blinder, Quantile regression

Table of Contents

1. Introduction

“It’s not a myth; it’s a math[1]”

The gender wage gap has been denied, it has been called a myth, been the centre of many discussions and literature and several solutions have been attempted in order to solve it. The gap comprises the differences in hourly wages between males and females. From World War II onwards when females started taking jobs in war industries, one of the first rules on this topic arose in the United States. The National War Labour board obliged companies in 1942 to equalize females’ salaries with males’ when executing work of the same quantity and quality or equal operational tasks. Employers did not comply with this voluntary request and job vacancies those days were even sex specific. For vacancies where no sex was mentioned, there would be separate pay scales for females and males. In general, for the years 1950-1960 a woman’s salary consisted of only 60% of their male counterparts. In 1963, one of the first anti-discrimination laws concerning the gender wage gap, The Equal Pay Act, was signed by President John F. Kennedy. Herewith the gender wage gap was officially acknowledged and addressed. Now 50 years later, the problem has still not been solved. In 2013 in Europe, the average gender pay gap was estimated at 16% (measured as the difference between male’s and female’s gross aggregate hourly earnings, as a percentage of male gross earnings). Figure A.1 in the Appendix shows the wage gaps in the year 2014 graphically. In Europe, Germany with a gap of 22% has one of the highest wage gaps of the continent. “This is unacceptable (…), Germany is one of the most economically developed countries and should lead by example, instead of lagging behind” EU commissioner Viviane Reding stated in the German newspaper Die Welt[2]. Striking is that such a high economically developed country, which even has a female leading the country for over ten years[3], has such a large gender wage gap. Germany is a country where they try to fight the gender wage gap with many structural labour market reforms amongst which the Federal Equality Law in 2001, the General Equal Treatment Act in 2006, and several political quota’s[4] (Directorate general for internal policies, 2013), but where the gender difference has barely shrunk over the last 15 years. These striking facts and the large struggle Germany seems to have with the gender wage gap in their country, interested me to investigate the wage gap in Germany specifically.

Finke (2011) of the Federal Bureau of Statistics found that the pay disparity between females and males could be due to multiple factors. Various literature (e.g Arulampalan et al., 2007; Blau and Kahn, 2016; Malmberg, 2007; Heinze, 2006; Montenegro, 2001; Gosse & Ganesh, 2002; Tansel, 2012) studied these multiple factors, and found industry factors, individual and family characteristics, experience and education as the main explanatory variables for pay disparities between genders. Only very few studies (Hsieh et al., 2012; Gambin, 2005) considered health as one of the possible contributing factors. This whilst health conditions are of great importance for human capital accumulation (Bartel and Taubman, 1979; Munshkin, 1962), which on its turn affects labour outcomes. Additionally there exists another gap between sexes, namely the gender health gap (Denton et al., 2004). Females are found to report on average lower health status than males. The question that arises subsequently, is whether these health differences contribute to the wage gap.

[1]Quote by Barack Obama (2014, April 4), in a speech about equal pay for equal work.
[2] Harman, S. (2010, March 5). Gender wage gap in Germany is among EU’s highest. *Die Welt*, Retrieved from http://www.dw.com.
[3] Since 22 November 2005, Angela Merkel (leader of the CDU since 2000) has been the chancellor of Germany.
[4] The Federal Equity Law, effective since 2001 aims at implantation of gender equity in the federal sector. The General Equal Treatment Act (AGG) became effective in 2006 and implements four European directives: the Racial Equality Directive, the Employment Equality Directive, the Gender Equality Directive for goods and services and the Employment Gender Equality Directive. One of its goals is to protect citizens from gender discrimination in employment. Quotas are implemented in German politics since 2013. The green party inserted a far reaching women’s statute in its articles and applies a quota of 50 percent, Die Linke also opted for a 50 percent quota and the Social Democratic Party 40 percent.

In this paper, two techniques will be applied for exploring the contribution of health to wages. Firstly, a Quantile regression analysis will be deployed, something that thus far has only been done once (Hsieh et al., 2012) when using health as an explanatory variable of wage. Quantile regression is a technique that is applied frequently in econometric research on the gender wage gap (e.g. Melly, 2006; Heinze, 2006; Montenegro, 2001; Tansel and Bodur, 2012). It allows estimation of the differential effects of the explanatory variables over the conditional distribution of the dependent variable. In this case, applying quantile regression allows to see the differences of the effects within several income quantiles. This is of great interest as general wage gap literature shows that the effects of explanatory variables such as experience and education on wages are different within income quantiles (e.g. Arulampam, 2007; Blau and Kahn, 2016; Albrecht et al., 2009 ; Machado and Mata, 2001). Subsequently, a decomposition method will be applied. This technique was first drafted by Oaxaca (1973) and Blinder (1973) and allows to separate the wage gap into two parts. The first part constitutes the differences in distributions of characteristics, such as health or education, among males and females, also named the 'explained' part of the wage differential. The second part represents the differences between sexes in returns to the explanatory variables and might show the actual discrimination. This part is called the 'unexplained' part of the wage differential, and will be the central issue in this paper. When looking at the contribution of health factors to the wage gap, the focus lies on the contribution of health factors to the 'unexplained' part of the wage gap, as literature (Denton et al., 2004) already proofs a contribution of health in the 'explained' part of the wage gap, as it shows a difference in distribution of health amongst genders.

How does health contribute to the gender wage gap by having different contributions of health to wages for males and females? This research question will be answered in this paper, by using data on German households from the German Socio-Economic Panel (GSOEP). For policy makers this is a very relevant question as there is still a big dilemma regarding the wage gap. If indeed health differences have a significant contribution to the wage gap, policy makers could whilom try addressing the gap from this new health perspective. The aim of this research paper is to find the "math" behind the gender wage gap, as Obama (2014) named it, and to formulate and quantify it. This will be executed starting with literature review on the topic and the development of a conceptual model in Section 2. Subsequently, in Section 3 the research methods and data will be described, and in Section 4 the resulting findings will be shown and discussed. In Section 5 limitations will be discussed and Section 6 concludes.

2. Theoretical framework

2.1 Literature review

As a first step in answering how health contributes to the gender wage gap by differing contributions of health on wages for females and males, I start with a literature review which provides me first information on this topic.

Effect of health on wages

The approach for investigating the wage gap was derived from human capital theory (Mincer, 1958; Becker, 1964). This theory states that the wage rate of an individual reflects its potential productivity, which is the result of its human capital characteristics. The importance of health as being part of this human capital was first emphasized by Becker (1964) and Fuchs (1966). Subsequently, Grossman (1972) was the first to motivate studying the relation between health and earnings, by including health in its model as a human capital component in which individuals will invest with the goal to increase possible working hours and productivity. From then, the finding that health is a part of an individual's human capital and through this it has an effect on labour market productivity, became more widely spread and nowadays is used in various research papers (e.g. Jäckle and Himmler, 2010; Contoyannis and Rice, 2001). When investigating this relationship, Leigh et al. (2009) indicate three potential mechanisms how health might affect economic inequality: labour market effects, marriage market effects and educational effects. Labour market effects explain the health-income relation through the hypothesis that poor health makes it harder for individuals to look for jobs, mentally and physically it will be more costly to work, and employers will less likely be hiring these individuals. Moreover, poor health negatively affects performance at the job or even working hours, which on its turn negatively affects income, job assurance and promotion chances. Marriage market effects, still a less strongly proven hypothesis, tries explaining a reduction in wage inequalities by the event of an adverse health shock to the population, which reduces the marriage rates (as healthy people are more likely to marry) and thereby alters the level of household income inequality. Lastly, educational effects show how during childhood poor health has an effect on school results due to concentration problems or sick absence, which in the end can result in lower earnings.

Differences between genders

One of the earliest papers distinguishing between genders in the health-earnings relation was Luft (1975). He found that there are differing impacts of poor health on earnings for genders, and in his study he concluded that these impacts where higher for females than for males. Interesting herewith is to shortly investigate how health is distributed amongst genders.

Even though females on average have lower mortality rates than males, they do report a greater variety and higher levels of health problems. As Denton et al. (2004) report, the roots of health inequalities are found in genetic, biological and social factors. Two main hypotheses are posed for explaining health inequalities between genders. First, the differential hypothesis, explains the inequalities by the reduced access of females to social and material conditions that increase health and the higher stress levels that are associated with their marital roles and gender. The second one, the differential vulnerability hypothesis, finds the reason for health inequalities in differences in reactions to psychosocial, behavioural and material conditions. Subsequently, higher health problems of females are also attributed to their exposure to higher obligation levels of society, more stressful life events and lower self-esteem (Denton et al., 2004).

After Luft (1975), few other studies started separating health effects on wage by gender. Thomas and Strauss (1997) used height, BMI and calories intake as health indicators and found that height positively affects wages for both males and females. Moreover, BMI only results in higher wages for males, and lower levels of per capita calorie intakes reduces wages of market workers. Contoyannis and Rice (2000) find that excellent self-assessed health positively affects female’s hourly wage and reduced psychological health reduces the hourly wage of males. Lastly, Gambin (2005) notes that the contribution of health to wage has been studied only limited and the divergence between genders herewith is even rarer. She finds that self-assessed health stronger affects male wages than those of females. The chronic illness indicator however is more significant for females.

Differences along the income distributions

When distinguishing between different income levels in investigating gender wage gaps, Arulampalam et al. (2007) found that gender wage gaps are typically larger at the higher income levels, so at the top of the wage distribution. This finding coincides with the suspicions of glass ceilings. Sometimes there is also a case of so-called ‘sticky floors’, which define (smaller) wage gaps at the bottom of the wage distribution. Differences in returns are one of the explanations for variation in wages across all income levels (Albrecht et al., 2003).

Resulting from this literature study, is a subdivision of the main question of this research into smaller questions. The first sub question is how health affects wages in the sample used for this research, especially divided per gender. The second sub question holds how the contribution of health to wage differs over income quantiles for this sample. The last question is focussed on how health has a contribution to the wage differential between genders.

2.2 Methods and conceptual model

The analysis in this paper will be based on the Mincerian wage function established by Mincer (1974) which is a human capital earnings function defining the natural logarithm of the wage rate as a function positively affected by schooling and experience:

$$\ln[W_i(s,p)] = \alpha_{it} + \gamma_{it}\, s + \beta 1_{it} p + \beta 2_{it} p^2 + \varepsilon \text{ (1)}$$

Where W_i is the natural logarithm of wage for individual *i, s* is the explanatory variable of schooling and *p* of experience. The use of a quadratic term is for explaining (Heckman and Polacheck, 1974) the existence of a concave relation between earnings and experience, due to the effect of a declining age-earning profile for a given level of experience (Montenegro, 2001). The model that I will use for this research will be a Mincerian wage function complemented with health variables and socio-demographic explanatory variables on which I will go into more detail? I will detail more ? in the data section. Regression by means of an ordinary least squares (OLS) model will allow me to get acquainted with the data, the variables and their relation. Such an OLS model is based on the mean of the conditional distribution of the dependent variable (Tansel and Bodur, 2012). It might however also be interesting to regress the effects of the explanatory variables at different quantiles of the dependent variable:

'"On the average" has never been a satisfactory statement with which to conclude a study on heterogeneous populations. Characterisation of the conditional mean constitutes only a limited aspect of possibly more extensive changes involving the entire distribution.'
Buchinsky (1994, page 453).

The Quantile regression technique allows to estimate impacts within different quantiles of the income distribution across the sample (McGuinness and Bennett, 2007). The Quantile regression model, first introduced by Koenker and Basset (1978) and later by Buchinsky (1994) looks as follows:

$$\ln W_i = \beta_\theta X_i + \varepsilon_{\theta i} \text{ with } Quant_\theta(lnW_i|X_i) = \beta_\theta X_i \text{ (2)}$$

Where $\ln W_i$ denotes the logarithm of wage of an individual, β_θ entails the vector of parameters and X_i the vector of explanatory variables for the given individual *i*. On this first part of the formula is the Quantile regression based. $Quant_\theta(lnW_i|X_i)$ denotes the θth conditional quantile of the dependent variable $\ln W_i$ given the explanatory variables X_i. The θth regression quantile, 0 < θ < 1 is defined as a solution for the following minimization problem:

$$\sum \rho_\theta(lnW_i, X_i|\beta_\theta) \text{ or } \sum \rho_\theta(\ln W - X_i\beta_\theta) \text{ (3)}$$

Where ρ_θ, which denotes the so-called 'checkpoint function', is defined as

$$\rho_\theta(z) = \theta z \; if \; z \geq 0 \; or \; \rho_\theta(z) = (\theta - 1)z \; if \; z < 0 \text{ (4)}$$

This Quantile regression technique does not minimize the sum of the squared residuals (as done in OLS regression) but it minimizes an asymmetrically weighted sum of absolute errors (Tansel and Bodur, 2012; Koenker and Hallock, 2001). Applying this Quantile Regression technique to my data allows to see the effects of the health variables on wage on the different income quantiles.

Subsequently, I will decompose the wage gap of this sample. The decomposition of gender wage gaps through the use of a linear regression framework was first drafted by Oaxaca (1973) and Blinder (1973). The framework estimates log-linear wage regressions by using subsamples of females and males. The variations in the coefficients estimates, which are multiplied by a group of characteristics, are due to the wage differential for an individual possessing this specific characteristics. This decomposition technique allows to disentangle the contribution of health on gender inequalities in wages in a part due to differences in the distribution of characteristics (explained differential) and a part due to differences in rewards to these characteristics (unexplained differential). The latter is sometimes mentioned as an estimation of gender discrimination (Blau and Kahn, 2016). The Oaxaca-Blinder decomposition is depicted by the following equations. For individual *i* in year *t*, the OLS wage regressions for female (F) and male (M) separately (subscripts *i* and *t* are eliminated for simplicity reasons) based on the earlier addressed Mincerian function (1):

$$\ln W_F = \beta_F X_F + \varepsilon_F \quad (5)$$
$$\ln W_M = \beta_M X_M + \varepsilon_M \quad (6)$$

Where *ln W* is the natural logarithm of hourly wage, *X* the vector of explanatory variables such as education, experience, health, β the vector of the corresponding coefficients and *ε* the error term.

Now, let b_F and b_M denote the OLS estimates of the corresponding β_F and β_M, and a bar over a variable designates mean values. As the residuals produced by OLS with a constant term have a zero mean, I can write:

$$\underbrace{\overline{lnW_M} - \overline{lnW_F}}_{(a)} = \underbrace{b_M\overline{X_M} - b_F\overline{X_F}}_{(b)} = \underbrace{b_M(\overline{X_M} - \overline{X_F})}_{(c)} + \underbrace{\overline{X_F}(b_M - b_F)}_{(d)} \quad (7)$$

Where (a) depicts the difference in the natural logarithm of wages between males and females; the gender wage gap. This gap is decomposed in (c) the impact of gender differences in endowments of explanatory variables also named the explained differential (Blau and Kahn, 2016), and in (d) the differential that is unexplained. The unexplained part denotes the difference in wages when females would have the same characteristics as males, and thus shows the differences in returns to these characteristics between genders. This might render a form of gender discrimination.

A generalization of the Oaxaca-Blinder decomposition technique is the Machado-Mata method (Machado and Mata, 2001; Machado and Mata, 2005). This technique can decompose the wage gap using Quantile regression and thus allows to show a decomposition of the wage differential in an explained and an unexplained part at different points in the wage distribution. An important limitation of this technique however is that it does not allow for detailed decompositions to be computed, which would allow computing the effect of all explanatory variables on the unconditional quantile wage distribution. As this research requires a detailed decomposition in order to find the effect on wage differentials of specific variables, I choose to use the Oaxaca-Blinder technique instead of the Machado-Mata decomposition technique for this research.

3. Research methods

3.1 Data

The study will be executed by use of the German Socio-Economic Panel Study (GSOEP) made available by the German Institute for Economic Research (DIW), Berlin. The GSOEP is a panel data set on the German population, assembled on a yearly basis which started in 1984. It primarily aims to collect information on the micro level of individuals, families and households, focussing on social and economic behaviour. For this research specifically, I created a sample consisting of 20 waves, ranging from 1994 to the year 2013. The years 1984-1993 were dropped from the sample due to missing observations in crucial variables and the years 2014-2015 were not yet available in the dataset. Subsequently, as I will be researching wage as dependent variable I will confine the sample to individuals in paid employment. Self-employed and unemployed individuals were therefore deleted from the sample together with observations with missing information. Lastly, individuals working less than 16 hours a week and individuals above the age of 65 were also deleted in order to surely have the active workforce. The final data sample used for this research consists of 20 waves of a year each and 149,019 observations. Table A.1 in the Appendix provides an overview of all variables including descriptive statistics.

3.2 Variables

Dependent variable

For the dependent variable the natural logarithm of hourly wage is used. As there was no direct variable available for hourly wage in the dataset, information on agreed weekly working hours is combined with gross monthly wage for obtaining the right information. The hourly wage is given in euro's and the averages are shown for males and females in Table 1 below, complemented with the calculated wage gap of this sample. In this sample the average hourly wage of females consists of 79% of the average hourly wage of males, a total gap of €3.81. This gap is graphically depicted in the Kernel density function in Figure 1 below.

Male Mean Wage	Female Mean Wage	Gender Gap (Wm-Wf)	Gender Gap As % of Wm	Gender Gap As % of Wf	Wf as % of Wm
€18.29	€14.48	€3.81	20.83%	26.31%	79.17%

Table 1 - Hourly wage statistics of male versus female

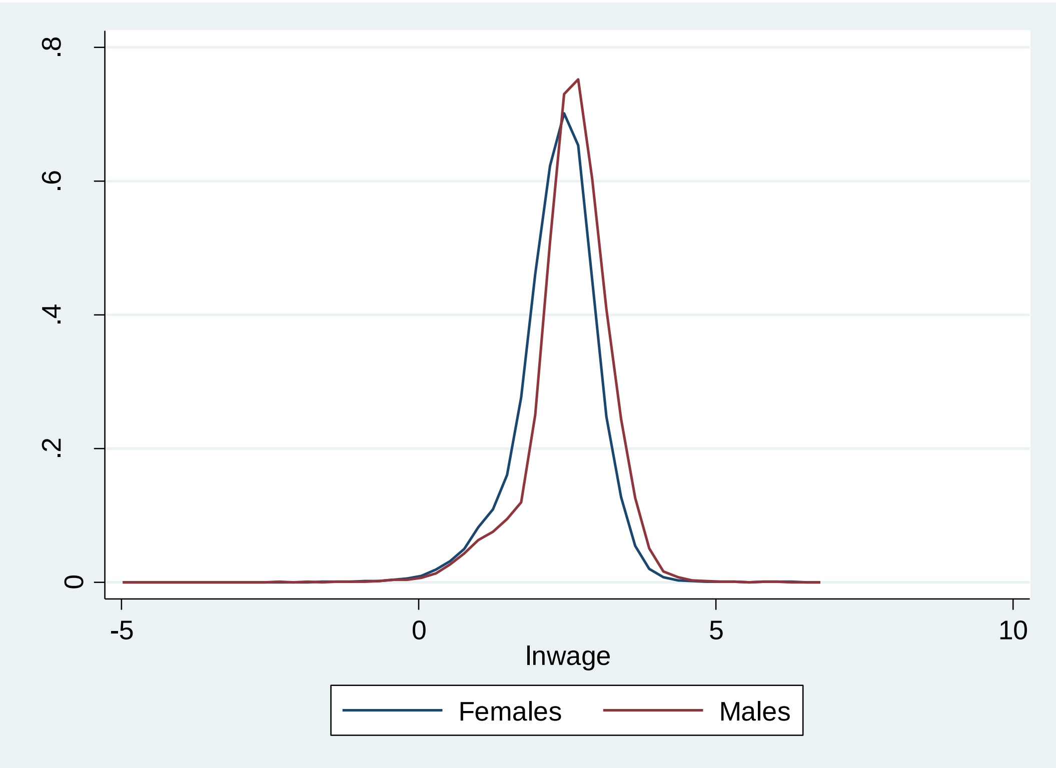


Figure 1 – Kernel density function of wages. A Kernel density function depicts the probability density of a chosen variable. In this figure it shows that males have a higher probability density for having a higher wage, than females.

Explanatory variables

In order to try to explain this gap, this paper looks at two types of health explanatory variables (also found in Table A.1 in the Appendix). First actual health impairment dummies that equal one whenever the individual has/ has had a health impairment such as a stroke, diabetes, cancer, psychiatric problems, angina or heart condition, or an attested disability or more than 30%. These are all the health impairment variables that are available in the dataset, and indicate actual, diagnosed health problems. This makes them an objective measure of health. The second type of health variable, is the self-rated health status which indicates each person's self-rated health status and is thus a subjective measure of health. This variable has the values 1 (very good), 2 (good), 3 (satisfactory), 4 (poor) and 5 (bad), and is converted to dummies for each value. When looking at the distribution of the self-assessed health values in Table A.1, it shows that females on average have a lower self-assessed health as their answers lie more in the lower range than it does for males. For the males the percentages are greater in the higher values. The mean of each health dummy in Table A.1 show that the actual health impairments are quite evenly distributed amongst genders.

For the non-health variables, I choose to include socio-demographic variables including age, education, working experience, employment level, marital status, and children in the household, based on precedent wage gap research (e.g. Arulampalan et al, 2007; Beblo and Berninger, 2003; Blau and Kahn, 2016). The average age of males (41.8) is almost equal to the average age of females (41.4) in this sample. The females in this sample have on average less children aged 0-14 in their household (0.37) compared to the males (0.58), but the level of children aged 15-18 is almost equal (0.19) for both. When looking at education, females have had slightly more years of education (12.4) than the males (12.3), but males have had more years of working experience in full time jobs (19.3 compared to 13.2). In current employment levels we see that 89% of the males and only 59% of the females, have a fulltime job. In order to control for time-effects, I included a dummy for each wave of the survey. Lastly, a dummy for the industry where the individual works is included, based on the findings of Heinze (2006) that not only human characteristics, but also industry characteristics should be taken into account when investigating such a wage gap.

3.3 Model specifications

Prior to continuing with the results, few specifications on the model will be discussed in this section. First, all health explanatory variables will enter the model as dummy variables. The impairment dummies will equal one when the individual has/had a health impairment and the self-assessed health variables are subdivided over five dummies, for each level one. Years of education and working experience enter the model as continuous variables, and this is also true for the variables age and number of children in the household. The variables marital status, waves and industry are split into dummies for each level. Second, In the decomposition of the model, normalized decomposition will be applied to all categorical regressors. This is to account for the fact that detailed decomposition results depend on the choice of the (omitted) base category. A solution for this problem is to decompose based on normalized effects. These are effects which are expressed as a deviation in contrast with the grand mean (Yun, 2005). Third, for the OLS regression I applied the use of robust standard errors, in order to correct for possibilities of heteroscedasticity. Heteroscedasticity, the case where the variability of a variable across the range of values of the variable predicting it, is unequal, results in inefficient OLS estimates with incorrect OLS standard errors. Applying robust standard errors corrects for this problem. Last, important to note is that the results cannot be interpreted as causal effects of health on wages. As will be explained more extensively in the discussion section, due to possible endogeneity issues in this research, interpretation of the results must be done carefully. Therefore, during this paper I will be referring to the results as associations between the explanatory variables and wages, and I will explain the contributions of these variables to the gaps rather than naming them as effects on the gap.

4. Results

4.1 OLS and Quantile regression

In this section I will present the OLS and the Quantile regression results of the explanatory variables on the natural logarithm of wage separated for females and males. The OLS regression will provide insights in the contributions of the explanatory variables to wage, and the differences between genders herewith. Subsequently, the Quantile regression will give a more thorough insight as it explains differences of the contributions of the explanatory variables to the logarithm wage quantiles 10%, 25%, 50%, 75%, 90%, instead of just investigating means as in OLS regression.

The first step in the empirical analysis of the gender wage differential is estimation of the log wage equations for males and females by applying OLS. Table 2 (males) and Table 3 (females) below show the OLS coefficients accompanied by their standard errors and the Quantile regression results.

MALES, on lnwage	**OLS Regression**		**Quantile Regression**				
Variables	Coefficient	Robust Std. Errors	θ=0.1	θ=0.25	θ=0.5	θ=0.75	θ=0.9
HEALTH VARIABLES							
Stroke	-.129		-.138	-.116	-.191	-.188	-.138
Diabetes	.033		.012	.004	.031	.038	.049
Cancer	.063		.069	.029	.061	.016	.048
Psych	-.044		-.042	.004	-.026	-.027	-.052
Heart	.004		-.029	.036	.026	.019	-.001
Disabl	-.041		-.017	.021	.018	-.016	-.031
Sahvg	.160		.181	.145	.126	.115	.147
Sahg	.116		.131	.105	.091	.079	.093
Sahsat	.070		.078	.059	.056	.044	.055
Sahpoor	.078		.110	.074	.056	.042	.049
Sahbad (ref)	-		-	-	-	-	-
OTHER VARIABLES							
Yearseduc	.080	.001	.071	.072	.077	.079	.079
Expft	.017	.001	.007	.008	.018	.019	.018
Sqexpft	-.000	.000	.001	.000	-.000	-.001	-.000
Exppt	-.012	.002	-.021	-.005	-.004	-.007	-.012
Sqexppt	.001	.000	.001	.000	.000	.000	.001
Age	.046	.001	.102	.0707	.024	.014	.010
Agesq	-.000	.000	-.001	-.001	-.000	-.000	-0.000
C014	.026	.002	.042	.029	.027	.024	.022
C1518	-.043	.003	-.061	-.054	-.019	-.012	-.014
Mmarried (ref)	-		-	-	-	-	-
Msingle	-.096		-.131	-.089	-.068	-.061	-.065
Mwidowed	-.076		-.084	-.096	-.067	-.039	-.051
Mdivorced	-.067		-.065	-.071	-.052	-.059	-.049
Mseparated	-.026		.013	-.020	-.018	-.033	-.033
Fulltime	.162		.304	.236	.116	.058	.061
Parttime (ref)	-		-	-	-	-	-
W1 (ref)	-		-	-	-	-	-
W2	.038		.041	.042	.029	.030	.025
W3	.081		.080	.077	.080	.078	.069
W4	.091		.119	.087	.089	.090	.072
W5	.090		.087	.092	.089	.109	.092
W6	.101		.104	.104	.096	.111	.099
W7	.175		.221	.194	.159	.150	.135
W8	.191		.242	.205	.184	.171	.143
W9	.247		.286	.241	.227	.235	.249
W10	.276		.304	.278	.257	.270	.280
W11	.278		.319	.285	.257	.264	.281

W12	.280		.329	.287	.269	.270	.262
W13	.268		.309	.288	.263	.266	.269
W14	.264		.308	.275	.251	.264	.261
W15	.278		.318	.291	.259	.271	.274
W16	.278		.315	.290	.269	.281	.276
W17	.288		.311	.305	.281	.294	.290
W18	.308		.337	.322	.291	.296	.302
W19	.325		.363	.337	.304	.322	.315
W20	.341		.370	.341	.324	.342	.336
Industry1	-		-	-	-	-	-
Industry2	-.330		-.205	-.332	-.353	-.278	-.277
Industry3	.142		.302	.230	.173	.128	.076
Industry4	.066		.266	.135	.050	.039	-.008
Industry5	.100		.171	.153	.124	.117	.097
Industry6	.071		.144	.110	.096	.088	.066
Industry7	-.070		-.007	-.025	-.042	-.056	-.061
Industry8	.003		.122	.062	.013	-.003	-.026
Industry9	.229		.441	.332	.233	.213	.192
Industry10	0.059	0.008	.183	.111	.026	-.012	-.047
Cons	.087	.034	-1.539	-.558	.616	1.052	1.306

Table 2 – OLS and Quantile regression results for males
Note: i) All results are statistically significant at 5% except the results in grey.

FEMALES, on lnwage	**OLS Regression**		**Quantile Regression**				
Variables	Coefficient	Robust Std. Errors	θ=0.1	θ=0.25	θ=0.5	θ=0.75	θ=0.9
HEALTH VARIABLES							
Stroke	-.055		.104	-.008	-.042	-.080	-.200
Diabetes	-.024		-.087	-.066	-.013	.011	.029
Cancer	.076		.070	.082	.044	.025	.034
Psych	-.030		-.059	-.005	-.014	-.025	-.075
Heart	-.038		-.015	-.019	-.039	-.053	-.059
Disabl	-.039		.000	.030	.005	-.013	-.023
Sahvg	.096		.114	.118	.076	.061	.055
Sahg	.062		.077	.069	.040	.043	.039
Sahsat	.038		.040	.043	.028	.026	.017
Sahpoor	.037		.047	.055	.031	.029	.016
Sahbad (ref)	-	-	-	-	-	-	-
OTHER VARIABLES							
Yearseduc	.078	.001	.078	.073	.071	.076	.081
Expft	.016	.001	.018	.015	.013	.015	.018
Sqexpft	-.000	.000	-.000	-.000	-.000	-.000	-.000
Exppt	.007	.001	.011	.011	.008	.004	.002
Sqexppt	.000	.000	.000	.000	.000	.000	.000
Age	.060	.001	.102	.072	.048	.033	.033
Agesq	-.001	.000	-.001	-.001	-.001	-.000	-.000
C014	.004	.003	.006	-.004	.000	.013	.016
C1518	-.060	.004	-.092	-.091	-.048	-.028	-.017
Mmarried (ref)	-	-	-	-	-	-	-
Msingle	-.010		.013	.072	.003	-.002	.006
Mwidowed	-.013		.028	-.001	-.020	-.018	-.011
Mdivorced	.003		.006	-.003	.005	.002	-.012
Mseparated	.019		-.005	-.091	.012	.021	.033
Fulltime	.052		.120	0.082	.036	.003	-.014
Parttime (ref)	-	-	-	-	-	-	-
W1 (ref)	-	-	-	-	-	-	-
W2	.033		.033	.044	.036	.029	.040
W3	.082		.090	.110	.080	.081	.077
W4	.098		.120	.119	.097	.088	.092
W5	.107		.128	.140	.104	.103	.107
W6	.119		.112	.136	.117	.117	.125

W7	.157		.182	.190	.156	.148	.153
W8	.176		.176	.209	.181	.164	.173
W9	.233		.239	.259	.229	.223	.237
W10	.253		.265	.285	.248	.247	.259
W11	.253		.250	.282	.255	.252	.268
W12	.263		.268	.296	.268	.264	.276
W13	.255		.277	.292	.261	.260	.265
W14	.240		.231	.268	.246	.250	.266
W15	.257		.255	.281	.254	.257	.273
W16	.268		.258	.298	.270	.282	.313
W17	.280		.284	.298	.284	.295	.302
W18	.282		.282	.302	.279	.289	.318
W19	.299		.306	.316	.305	.296	.321
W20	.334		.365	.341	.330	.326	.356
Industry1 (ref)	-	-	-	-	-	-	-
Industry2	-.228		-.099	-.175	-.283	-.250	-.258
Industry3	.301		.437	.413	.273	.203	.140
Industry4	.376		.588	.326	.185	.144	.646
Industry5	.122		.200	.143	.082	.087	.078
Industry6	.186		.306	.206	.146	.146	.120
Industry7	-.016		.052	.024	-.049	-.075	-.087
Industry8	.117		.215	.185	.087	.054	.013
Industry9	.294		.452	.396	.266	.210	.165
Industry10	.110		.252	.194	.081	.042	.001
Cons	-.245		-1.748	-.704	.209	.675	.815

Table 3 – OLS and Quantile regression results for females

Note: i) All results are statistically significant at 5% except the results in grey.

Starting with the focus variables of this research, the health variables, the coefficients show that for males having (had) a stroke decreases wage with 12.9%, psychological problems decrease wage with 4.4%, and having (had) a disability negatively contributes to wage with 4.1%, all significant at 5%. These results are as, because earlier research showed that health problems negatively affect wages. Having (had) cancer increases wage with 6.3%, something that seems to be very strange and goes against gut-feelings. A possible explanation could be that the small population group (n=513) that has (had) cancer, by chance happen to have higher average incomes, which are not due to this disease. New research with a different sample would be required in order to see whether this indeed is the case. Lastly, diabetes and heart problems do not have a significant contribution to wages. The variable self-assessed health positively contributes to wage and the higher the self-assessed health, the higher the wage, all effects being significant. For example, having a good self-assessed health increases wage with 11.6% compared to having (had) a bad self-assessed health, and a very good self-assessed health increases wage with 16% compared to having (had) a bad self-assessed health. This is as expected, as it has been discussed in literature that the better individuals feel, the greater their productivity for example.

For females, only the health impairment variable disability significantly negatively contributes to wage by 3.9%, which is lower than the result for males (-4.1%). Again strangely, having (had) cancer significantly positively contributes to wages for females, in a higher manner (7.6%) than it does for males (6.3%). The other health impairment variables are all insignificant. This could mean that for this sample, female wages are less affected by health impairments than males. This might be a result from the type of work they do, or that for females, their job tasks suffer less from their health impairment than it does for males. The self-assessed health variables all positively contribute to wage in a significant manner. Same as for males, income increases with health satisfaction. For females however, this contribution is smaller than for males, as having a good self-assessed health increases wages with 6.2% compared to having a bad self-assessed health, and a very good health increases wages with 9.6% compared to a bad self-assessed health. For males, a very good self-assessed health has a stronger contribution (16%) to wages compared to having a bad health than for females (9.6%). Same as for the actual health impairments, it shows here that males are more strong affected by their health, both in a negative and positive way.

Continuing with the Quantile regression of the health variables, for males it depicts that most of the health impairment variables are insignificant over the quantiles. The variable for a stroke significantly negatively contributes to wage in the median (-19.1%) and the .75 quantile (-18.8%). Showing that this is only of importance in the higher income quantiles. Cancer is only significant at the median, where the contribution to wage is positive. The fact that this is only significant at the median might depict a (few) outlier(s) in this wage group, that happen to have a higher wage and the health impairment cancer. The fact that this variable is not significant over the rest of the income quantiles, weakens the OLS result of cancer having a positive contribution to wage. Lastly, having (had) a disability starts with a positive contribution to wages of 2.1% at the .25 quantile, decreasing to a 1.8% increase in wage due to this health impairment. Later, it starts to negatively contribute to wage from the .75 quantile on (-1.6%), decreasing even more in the .9 quantile to -3.1%. This decreasing trend shows that having (had) a disability becomes worse for wage, the higher the income quantile. So in the higher income quantiles, having (had) a disability has the greatest negative contribution to wage. This might be explained by disability being of a greater burden or problem for jobs that are in this higher income quantile. Self-assessed health contributes to income positively the strongest in the lowest income quantile and this strength reduces towards the higher income quantiles. For example, a very good self-assessed health increases income with 18.1% in the 0.1 income quantile, 14.5% in the .25 quantile, 12.6% in the .5 quantile and 11.5% in the .75 quantile, all compared to having a bad self-assessed health. In the .9 income quantile this contribution increases again to a 14.7% wage increase compared to having a bad self-assessed health. For all self-assessed health states, the income increase is higher in the lower income quantiles and lower in the higher income quantiles. Except for the .9 quantile, here the contribution of the self-assessed health states are stronger than in the .75 quantile. Possibly in the lower income quantiles, how the individual values his health has a greater impact on labor affecting factors such as productivity for example, than in the higher income quantiles. It could be that the individuals in the higher income quantiles are letting their self-assessed health affect their job performances less than individuals in the lower income quantiles.

For females, stroke (-20%), psychological problems (-7.5%), are only significant at the .9 income quantile. This shows that for the females in this sample, these health impairment variables are only a burden for the jobs that are in the higher income quantile. Having (had) a disability significantly positively contributes to wage with 3% at the .25 income quantile, and it significantly negatively contributes to wage with 2.3% in the highest income quantile. Following the same trend as for males. The self-assessed health variable also follows an almost equal trend as for the males, where the percentage increase is the highest in the lower quantiles and decreasing towards the higher quantiles. For the females however, in the .9 quantile there is no upward trend as it is for the males, but here the downward trend continuous. For example, for females having a very good self-assessed health positively contributes to income with 11.4% compared to having a bad self-assessed health, 11.8% in the .25 quantile, 7.6% in the .5 quantile, 6.1% in the .75 quantile and 5.5% in the .9 quantile. This is lower than for males, where this is 18.1%, 14.5%, 12.6%, 11.5%, and 14.7% respectively.

Now, I will shortly summarize the results of the non-health variables. For the male population in this OLS regression, all variables are significantly different from zero. Wage increases with years of education (8%), full time job experience (1.7%), age (4.6%) and children aged 0-14 years (2.6%). Being single, widowed, divorced or separated decreases wage compared to being married. For example, being single decreases wage with 9.6% compared to being married. This might be a result from happiness coming with marriage, or financially needing to care for the spouse. Working in the agriculture or trade industry reduces wage with 33% and 7% respectively, compared to other industries. It could be that in these industries average incomes are lower. Over the quantiles it shows that the size of the contributions of education and experience increase with income from 7.1% to 7.9% and 0.7% to 1.8% respectively. This could mean that education and experience becomes more and more important, the higher the income scale of the job is. The sizes of the contributions of age (10.2% to 1%), children between 0-14 years (4.2% to 2.2%), children between 15-18 years (-6.1% to -1.4%), and marriage decrease with income. Being single, for example, decreases wages with 13.1% in the .1 income quantile and only 6.5% in the .9 income quantile, compared to being married. This could be because males in the higher income quantiles are less sensitive for these marital status changes than the males in the lower income scale.

For females, wage also increases with years of education and full time job experience at almost equal sizes as for males. Age however, positively contributes to wage with only 0.6% for females. Having children aged 0-14 years has no significant contribution in contrast to males. Moreover, in contrast to males, for females only being single reduces wage with 1% compared to being married. The other marital statuses are insignificant, which could mean that females are not influenced by marital status as much. Working at the agriculture or trade industry also reduces wages, same as for males, but in lower sizes (22.8% and 1.6% respectively). Looking at the Quantile regression results, it shows that the variables years of education, full time experience, age, and children aged 15-18 follow an equal trend with comparable sizes as for males. Having children aged 0-14 for females only significantly positively contributes to wage at the .75 and .9 income quantile, sized 1.3% and 1.6% respectively. In contrast to the male results, for females marital statuses are insignificant at almost all quantiles. This could mean that concerning income, the females in this sample are not affected by changes in marital statuses in contrast to males.

4.2 Oaxaca-Blinder decomposition

Table 4 below presents the results of the Oaxaca-Blinder decomposition. In order to see the effect of adding health variables to the decomposition, the results are separated in a model without health variables, and a model with health variables added. The table shows the measurement of the wage gap using the log wage unit.

	Without health		**With health variables**	
Lnwage	**Absolute**	**Percent**	**Absolute**	**Percent**
Predicted mean males	2.774		2.774	
Predicted mean females	2.548		2.548	
Difference	.226		.226	
Decomposition into				
- **Explained**	0.071	31.4%	.072	31.9%
- **Unexplained**	0.155	68.6%	.154	68.1%
EXPLAINED	**Absolute**	**Percentage of total**	**Absolute**	**Percentage of total**
yearseduc	-.012	-16.8%	-.012	-16.7%
Exp	.029	40.8%	.029	40.2%
Age	.001	1.4%	.001	1.4%
Children	.008	11.3%	.008	11.1%
Emplstat	.030	42.2%	.030	41.7%
Wave	-.007	-9.9%	-.007	-9.7%
Industry	.022	31.0%	.022	30.6%
stroke			-.000	0%
diabetes			.000	0%
cancer			-.000	0%
pshyc			.000	0%
heart			-.000	0%
disabl			-.000	0%
SAH			.001	1.4%
Total	0.071	100%	0.072	100%
UNEXPLAINED	**Absolute**	**Percentage of total**	**Absolute**	**Percentage of total**
yearseduc	.022	14.2%	.018	11.2%
Exp	.009	5.8%	.010	6.1%
Age	-.200	-129.0%	-.173	-112.3%
Children	.020	12.9%	.020	13.0%
Emplstat	.031	20.0%	.031	20.1%

Wave	.001	0.6%	.001	0.6%
Industry	-.025	-16.1%	.024	15.4%
stroke			-.0001	0%
diabetes			.0001	0%
cancer			-.0001	0%
pshyc			-.0001	0%
heart			.0002	0%
disabl			-.0002	0%
SAH			.010	6.5%
_cons	.297	191.6%	.215	139.6%
Total	0.155	100%	0.154	100%

Table 4 – Oaxaca-Blinder decomposition

Notes:

i. Exp: expft sqexpft exppt sqexppt
ii. Age: age agesq
iii. Children: c014 c1518
iv. Emplstat: fulltime parttime
v. Wave: w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16 w17 w18 w19 w20
vi. Industry: industry1 industry2 industry3 industry4 industry5 industry6 industry7 industry8 industry9 industry10
vii. SAH: sahvg sahg sahsat sahpoor sahbad

All results are statistically significant at 5% except the results in grey.

In this sample, the average log wage for males is 2.774 and for females 2.548. This results in an average male-female wage differential of 0.226 log points, which is statistically significant. Subsequently, the wage differential is subdivided in an explained and an unexplained part. The explained part, of which the differential can be explained by differences in productivity characteristics between the two genders, equals 31.4% of the total wage gap in the model without health variables, and 31.9% for the model with health included. The unexplained part, of which the differences designate differences in returns to characteristics and thus which might imply discrimination, equals 68.8% of the wage gap of the model without health variables, and 68.1% points for the model with health variables. The major part of the wage differential is thus due to unexplained reasons. In order to find the contribution of the health variables to this unexplained differential, Table 4 above also entails the detailed Oaxaca-Blinder decomposition for both models.

Decomposition of the model without health variables

First the decomposition of the model without the health variables will be discussed here.

Explained differential

When looking at the detailed decomposition of the explained differential of the first model it shows that the major part of the explained wage differences are due to differences in experience (40.8%) and differences in employment status (42.2%) between males and females. As the explained differential depicts the wage differences due to differences in characteristics between the two genders, it can be concluded from this information that males have more experience and work more fulltime than females. Looking at Table A.1 in the Appendix, this indeed is true and therefore results in a wage gap in favor of males. Experience has a large contribution to the explained differential, because employers pay workers more, the more experience they have. This is because they might be more valuable to the employer due to this experience. Experience has such a large contribution in the explained differential specifically because differences in experience between genders logically leads to a difference in income between the two. For employment status equal logic accounts, as hourly wages can be expected to be higher for fulltime employees than for part-time employees, and thus account for an explainable gap in de wages when females work less fulltime and more part-time than males. The large contribution of waves (-9.9%) can be explained by the fact that the wave, or the year, in which the individual was interviewed, logically affects wages because for example wages where lower in the 70's compared to the 00's. That this value is negative for waves, meaning that it creates a wage gap in favor of females, could possibly explained by more females being interviewed in the later waves and males more in the earlier waves. The negative contribution of years of education (-16.8%) is due to the fact that females on average have more years of education. This difference in characteristics results in a difference in wages in favor to females.

Unexplained differential

Subsequently, I will look at the unexplained differential, the wage differential that cannot be explained by differences in characteristics between genders. This differential shows the differences in wages if females would possess the exact same characteristics as the males in the sample and thus indicates the actual wage gap. For this first model a major part is due to the variable age (-129.0%). This value indicates that females with the same age as males, would earn a higher wage. This is a strangely high result, and it could be that this is an omitted variable bias; which is a bias that is created in order to compensate for missing variables which are incorrectly left out of the model. I expect that in further research when other possibly important factors are added to the model, that this value will decrease. The variables that discriminate in favor of males are the variables children (12.9%) and employment status (20%), meaning that the wage returns to these characteristics are higher for males than for females. Years of education (14.2%) and experience (5.8%) also discriminate in favor of males, but these results are found insignificant.

This decomposition has a relatively large constant (0.297) of 191%. This means that a large portion of the gap is not explained by the gender differences of the chosen variables in the model. Including other relevant variables could reduce this constant, and thus this shows that further research with more relevant variables is required in order to reduce this constant, and also reduce the omitted variable bias found.

Decomposition of the model with health variables

In the second model, the health variables are added.

Explained differential

In the explained differential, the results are almost equal to the results in the model without the health variables. Again the greatest contributors to the explained differential are experience (40.2%) and employment status (41.7%). Education is here again a negative contributor (-16.7%) suggesting that females have more years of education on average, and therefore earn more, which is correct for this sample. When looking at the health variables that are added now, it shows that the variables stroke, diabetes, cancer, psychological problems, heart problems and disability do not have a distinctive effect (all 0%). The variable self-assessed health has a very small contribution (1.4%) to the explained wage differential, adverting to the small difference in self-assessed health values between males and females.

Unexplained differential

The detailed decomposition of the unexplained differential shows comparable results as for the model without the health variables, concerning the variables years of education, experience, age, children, employment status, and waves. Regarding the health variables, the decomposition shows for the health impairment variables stroke, diabetes, cancer, psychological problems, heart problems, and disability, that they have a neglectable size. The self-assessed health variable however forms 6.5% of the total unexplained wage differential, being statistically significant. This means that self-assessed health does contribute to the actual wage gap (the wage gap that cannot be explained by differences in distribution of characteristics between genders), however only with a relatively small percentage.

The constant of this model with health variables (0.215) is slightly lower compared to the model without health variables (0.297). This implies that adding the health variables explains the unexplained wage gap better than without adding them to the model. However, the model still entails a high constant. This means that there is still a large proportion of the gap that cannot be explained by the variables chosen in this model.

5. Discussion and conclusion

This section will start with possible limitations of this study being discussed, including their consequences for the results found in this study. Later, the main findings of this study will be presented and compared to the literature discussed, their implications for policy will be summarized, and recommendations will be given.

Starting with critically evaluating the results of this study, I will discuss the possibly limitations found in this study. First, the major problem in the regression used in this paper, is the expected endogeneity of the explanatory health variables. In the early days already many papers (Lee, 1982; Ettner, 1996; Stronks et al., 1997) have been written about the effect of income on health and shown a certain endogenous relation between the two variables. As many papers show (Duncan, et al., 2004; Blau and Kahn, 2016; Heinze, 2006; Jäckle, 2010), a possible result of this endogenous explanatory variable is that the regressions do not measure the causal effect consistently. This has to be taken into account when interpreting the results of this study.

Second, using self-assed health as an explanatory health variable is questionable as it does not assess one's actual health status and due to biases measurement errors can result. Self-assessed health predict also other outcomes which the individual can care about such as happiness for example. This problem is partly solved by adding objective health indicators. The objective health indicators in this research as stroke, diabetes, cancer, psychological problems, heart problems, and disability, might however be too specific and too severe. Possibly inclusion of smaller health problems such as back or neck problems, being sick often, having no energy, etcetera, might have an impact on the wage differential between genders, and thus might be interesting for further research. Moreover, it is likely that unobserved effects such as genetic endowments for example are correlated with the health variables. When a person is genetically 'better' equipped, he or she might be healthier and also receive a higher salary than an individual without this equipment. These unobserved effects might lead to a potential omitted variable bias which violates the particular assumption of the linear regression model that has an error term uncorrelated with the regressors. This results in the OLS estimator being biased and inconsistent.

Third, a disadvantage of choosing the Oaxaca-Blinder method is that this method only focusses on average effects, so only confined to the mean. This leads to incomplete or even misleading assessments of the effects of the explanatory variables when their effects differ across the distribution of the dependent variable. In case of the sample used for this research, the effects of the explanatory variables indeed differ across income quantiles, as can be seen in the Quantile regression in Table 2 and Table 3. There is a decomposition method available that decomposes wage gaps for example over the quantile distributions (Machado and Mata, 2001; Machado and Mata, 2005). However, this method does not allow one to look at the detailed decomposition of a differential, meaning that it does not show the contribution of each variable on its own to the whole differential. Due to this limitation I decided not to apply the Machado Mata decomposition technique, as the goal of this research required a detailed decomposition of the differential into the contribution of all variables specific in order to find the contribution of the health variables to the total wage gap.

Lastly, the high constant resulting from the Oaxaca-Blinder decomposition points at the fact that still a large part of the unexplained differential cannot be explained by the variables chosen in the model. The constant captures the gender differences in variables that are unobserved by the chosen model, and thus they will be picked up into the intercept. Further research thus would be required in order to find out which important variables are missed in this research. The article of Heinze (2006) states that firm characteristics such as firm results, quota's, technical state, wage agreement, sales, business start-up after a specific period, number of employees, etcetera have a high impact on wage differentials. As these variables where not included in the chosen data set they were not possible to be included in this particular research. For further research I would recommend to include a set of firm characteristics to the model, and then see how much health variables still contribute to the total wage differential. If these firm characteristics are indeed relevant explanatory variables the constant will be reduced by adding these to the model.

This paper has examined the gender wage gap in Germany using the German Socio-Economic Panel data set. Central to this research was the question how health contributes to the gender wage gap, subdivided in the sub questions how health affects wages per gender, how the effect of health on wage differs over income quantiles and finally how health impacts the wage differential between genders. This research was executed by applying OLS regression, Quantile regression and Oaxaca-Blinder decomposition techniques.

The results of this research confirm the earlier findings in literature of health having an influence on wages, where the finding of this research is that this contribution is larger for males than for females. This contradicts the findings of Luft (1975), who found that poor health influence earnings stronger for females than for males. Luft (1975) however only looked at whether individuals' reported "well" or "sick", which is a different measure for health than self-assessed health. The research results of this paper do confirm the findings of Gambin (2005) that self-assessed health stronger affects male wage than those of females. This effect differs over income quantiles, as also found by Arulampalam et al. (2007). It is stronger in the lower in the lower income quantiles and it reduces the higher the quantile. Moreover, this research shed a new light on this subject as it shows how health not only has an effect on wages, but it also has a small contribution to the wage differences between genders, in the unexplained wage differential specifically. Something that has not been found in other literature so far, and is thus a new contribution to literature. Even though objective health problems such as a stroke, diabetes, cancer, psychological problems, heart problems, and disability show no effect on the unexplained wage differential, the self-assessed health variable does with 6.5%. The effect is relatively small, but it is significantly different from zero. I would recommend to conduct further research on this topic with including even more health variables in order to see if indeed objective health problems have no effect on the unexplained wage differential, or whether only in this study they do not have an effect due to the specific chosen health problems.

For now, these results do not have strong implications for policy, as the contribution of health factors to the wage gap are relatively small. Moreover, trying to change the differences in self-assessed health between genders might be difficult. These factors imply a great deal of work with only a small result, something that might not interest policy makers that much. However, results from further research on this topic might be interesting for policy makers, as with this research larger contributors to the wage gap can be found. To conclude, the "math" behind the wage gap has been partially exposed in this research, also by inclusion of health variables, but there still is a big part that has to be identified yet.

6. References

Albrecht, J., Van Vuuren, A., & Vroman, S. (2004). Decomposing the gender wage gap in the Netherlands with sample selection adjustments.

Albrecht, J., Van Vuuren, A., & Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics*, *16*(4), 383-396.

Arulampalam, W., Booth, A. L., & Bryan, M. L. (2007). Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial & Labor Relations Review*, 60(2), 163-186.

Badel, A., & Peña, X. (2009). Decomposing the gender wage gap with sample selection adjustment: Evidence from Colombia. *Bogotá, Colombia: Universidad de las Andes. Mimeographed document. Available at: http://ximena. pena. googlepages. com/gender. pdf*.

Bartel, A., & Taubman, P. (1979). Health and labor market success: The role of various diseases. *The Review of Economics and Statistics*, 1-8.

Becker, Gary, Human Capital and the Personal Distribution of Income, Woytinsky Lecture No. 1, University of Michigan, 1967

Beninger, D., Heinze, A., & Laisney, F. (2003). *Measuring selectivity-corrected gender wage gaps in the EU*. Zentrum für Europ. Wirtschaftsforschung.

Blau, F. D., & Kahn, L. M. (2016). The Gender Wage Gap: Extent, Trends, and Explanations.

Blinder, A. (1973): Wage Discrimination: Reduced Form and Structural Estimates. J*ournal of Human Resources*, Vol. 8: 436-455.

Buchinsky, M. (1994) 'Changes in the US wage structure, 1963-1987: application of quantile regression', *Econometrica*, 62, 405-458

Cai, L., & Kalb, G. (2006). Health status and labour force participation: evidence from Australia. *Health economics*, *15*(3), 241-261.

Contoyannis, P., Jones, A. M., & Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, *19*(4), 473-503.

Contoyannis, P., & Rice, N. (2001). The impact of health on wages: evidence from the British Household Panel Survey. *Empirical Economics*, *26*(4), 599-622.

Currie, J., & Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of labor economics*, *3*, 3309-3416.

Davies, H., Joshi, H., & Peronaci, R. (2000). Forgone income and motherhood: What do recent British data tell u?. Population studies, 54(3), 293-305.

Denton, M., Prus, S., & Walters, V. (2004). Gender differences in health: a Canadian study of the psychosocial, structural and behavioural determinants of health. *Social science & medicine*, *58*(12), 2585-2600.

Directorate general for internal policies . (2013). Gender Equality: The recruitment practices in top management and non-executive directors' positions in Europe. *Workshop: Policy Department Citizen's Rights and Constitutional Affairs* , 1(1), . Retrieved 17 March, 2016, from http://www.europarl.europa.eu/RegData/etudes/workshop/join/2013/474410/IPOL-FEMM_AT(2013)474410_EN.pdf

Doyal, L. (2000). Gender equity in health: debates and dilemmas. *Social science & medicine*, *51*(6), 931-939.

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies. *Research in Human Development*, *1*(1-2), 59-80.

Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of health economics*, *15*(1), 67-85.

Eurostat (2016, March 1). Gender pay gap statistics. Retrieved from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics.

Finke, C. (2011). Verdienstunterschiede zwischen Männern und Frauen. Eine Ursachenanalyse auf Grundlage der Verdienststrukturerhebung 2006. *Wirtschaft und Statistik, H*, *1*, 36-50.

Fitzenberger, B., Schnabel, R., & Wunderlich, G. (2004). The gender gap in labor market participation and employment: A cohort analysis for West Germany. *Journal of population economics*, *17*(1), 83-116.

Fitzenberger, B., Koenker, R., & Machado, J. A. (Eds.). (2013). *Economic applications of quantile regression*. Springer Science & Business Media.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. *Handbook of labor economics, 4,* 1-102.

Gambin, L. M. (2005). The impact of health on wages in Europe–does gender matter. *HEDG WP*, *5*(03).

Gosse, M. A., & Ganesh, S. (2002). The gender pay gap in the New Zealand public service. *State Services Commission*.
Harman, S. (2010, March 5). Gender wage gap in Germany is among EU's highest. Die Deutsche Welle, Retrieved from http://www.dw.com.
Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*. *January, 47*(1), pp. 15361.
Heckman, J., & Polachek, S. (1974). Empirical evidence on the functional form of the earnings-schooling relationship. *Journal of the American Statistical Association*, *69*(346), 350-354.
Heinze, A. (2006). Decomposing the Gender Wage Differences using Quantile Regressions. *Centre for European Economic Research, ZEW Mannheim.*
Hsieh, W. J., Hsiao, P., & Lee, J. (2012). The impact of Health status on wages–Evidence from the quantile regression. *Journal of International and Global Economic Studies*, 1(5), 35-56.
Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.
Jones, F. L., and J. Kelley (1984). Decomposing differences between groups: A cautionary note on measuring discrimination. *Sociological Methods and Research 12*.
Jäckle, R., & Himmler, O. (2010). Health and Wages Panel data estimates considering selection and endogeneity. *Journal of Human Resources*, *45*(2), 364-406.
Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453-479.
Koenker, R. G. Basset (1978) "Regression Quantiles," *Econometrica*, 46(1): 33-50
Koenker, R. and K. Hallock (2001) "Quantile Regression," *Journal of Economic Perspectives,* 15(4): 143-156.
Kunze, A. (2008). Gender wage gap studies: consistency and decomposition. *Empirical Economics, 35*(1), 63-76.
Lee, L. F. (1982). Health and wage: A simultaneous equation model with multiple discrete indicators. *International Economic Review*, 199-221.
Leigh, A., Jencks, C., & Smeeding, T. M. (2009). Health and economic inequality. *The Oxford Handbook of Economic Inequality, Oxford University Press, Oxford*.
Luft, H. S. (1975). The impact of poor health on earnings. *The Review of Economics and Statistics*, 43-57.
Machado, J. A., & Mata, J. (2001). Earning functions in Portugal 1982–1994: Evidence from quantile regressions. *Empirical Economics*, *26*(1), 115-134.
Machado, J. A., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, *20*(4), 445-465.
Malmberg, Å. (2007). Evaluating the gender wage gap in Sweden.
Martins, P. S., & Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour economics*, *11*(3), 355-371.
Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. Review of Labor Economics, 68(4), 543-572.
McGuinness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. *Economics of Education Review*, *26*(5), 521-531.
Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings* (pp. 41-63). NBER.
Montenegro, C. (2001). *Wage distribution in Chile: Does gender matter? A quantile regression approach*. World Bank, Development Research Group/Poverty Reduction and Economic Management Network.
Mushkin, S. J. (1962). Health as an Investment. *The Journal of Political Economy*, 129-157.
Nicodemo, C. (2009). Gender pay gap and quantile regression in European families.
Nikolka, T. (2013). The Gender Wage Gap in OECD Countries. *DICE Report, 11*(1), 69.
Oaxaca, R. (1973): Male-Female Wage Differentials in Urban Labor Markets*, International Economic Review*, Vol. 14: 693-709.

Stronks, K., Van De Mheen, H., Van Den Bos, J., & Mackenbach, J. P. (1997). The interrelationship between income, health and employment status. *International Journal of Epidemiology*, *26*(3), 592-600.
Tansel, A., & Bodur, F. B. (2012). Wage inequality and returns to education in Turkey: A quantile regression analysis. *Review of Development Economics*, *16*(1), 107-121.
Van Doorslaer, E., & Gerdtham, U. G. (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social science & medicine*, *57*(9), 1621-1629.
Vogel, L., Robertson, G., & Altman, J. G. (2003). *The gender workplace health gap in Europe*. TUTB.
Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys, 19*(3), 479-511.
Yun, M. S. (2005). Normalized equation and decomposition analysis: computation and inference.
Zhang, H., d'Uva, T. B., & van Doorslaer, E. (2015). The gender health gap in China: A decomposition analysis. *Economics & Human Biology*, *18*, 13-26.

Appendix A – Tables

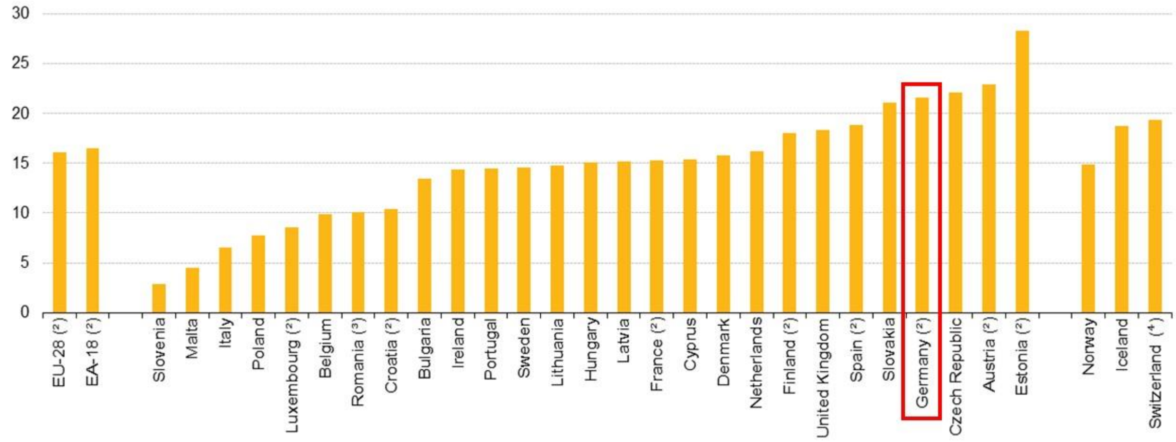
Table A.1 - Variable names, definitions and descriptive statistics.

VARIABLE	DEFINITION	MALES 82,323 obs		FEMALES 66,696 obs	
		MEAN	ST. DEV	MEAN	ST. DEV
DEPENDENT					
Lnwage	Natural logarithm of gross hourly wage in euro's Gross hourly wage is calculated by dividing gross income by four times the amount of agreed weekly working hours, of only people in dependent employment and who work >16hours a week.	2.774	.546	2.548	.533
HEALTH EXPL					
Diabetes	1 if individual has or had diabetes, 0 if not	.006		.004	.
Cancer	1 if individual has or had cancer, 0 if not	.002		.005	
Pshych	1 if individual has or had psychological problems, 0 if not	.005		.011	
Heart	1 if individual has or had heart problems, 0 if not	.006		.004	
Disabl	1 if individual has or had a disability, 0 if not	.056		.050	
Sah	Self-assessed health status (sah)				
Sahvg	Sah – 1 very good, 0 otherwise	.106		.099	
Sahg	Sahg – 1 good, 0 otherwise	.492		.472	
Sahsat	Sah – 1 satisfactory, 0 otherwise	.306		.314	
sahpoor	Sah – 1 poor, 0 otherwise	.084		.101	
Sahbad	Sah – 1 bad, 0 otherwise	.011		.014	
OTHER EXPL					
Age	Age of individual in years	41.8	11.0	41.4	11.0
Agesq	Agesquared/100	1865	924	1836	904
C014	Number of children aged 0-14 years in hh	.588	.902	.370	.684
C1518	Number of children aged 15-18 years in hh	.191	.458	.194	.456
Yearseduc	Years of education completed	12.296	2.690	12.446	2.591
Expft	Working experience fulltime – total length of fulltime employment so far	19.277	11.934	13.230	10.356
Exppt	Working experience parttime – total length of parttime employment so far	.439	1.652	4.488	6.483
Marital status	**Present marital status of individuals >16y**				
Mmarried	Marital status – 1 married, 0 otherwise	.667		.584	
Msingle	Marital status – 1 married, 0 otherwise	.249		.261	
Mwidowed	Marital status – 1 married, 0 otherwise	.005		.025	
Mdivorced	Marital status – 1 married, 0 otherwise	.061		.105	
Mseparated	Marital status – 1 married, 0 otherwise	.0181		.024	
Employment level	**Level of employment**				
Fulltime	Employment level – 1 fulltime, 0 otherwise	.888		.587	
Parttime	Employment level – 1 parttime, 0 otherwise	.111		.413	
Wave dummies	**Dummy for each wave**				
W1	1 wave 1994, 0 otherwise	.044		.039	
W2	1 wave 1995, 0 otherwise	.045		.040	
W3	1 wave 1996, 0 otherwise	.043		.039	
W4	1 wave 1997, 0 otherwise	.041		.037	
W5	1 wave 1998, 0 otherwise	.043		.039	
W6	1 wave 1999, 0 otherwise	.043		.034	
W7	1 wave 2000, 0 otherwise	.063		.058	
W8	1 wave 2001, 0 otherwise	.061		.057	
W9	1 wave 2002, 0 otherwise	.061		.060	
W10	1 wave 2003, 0 otherwise	.059		.059	
W11	1 wave 2004, 0 otherwise	.054		.055	
W12	1 wave 2005, 0 otherwise	.052		.052	
W13	1 wave 2006, 0 otherwise	.052		.054	
W14	1 wave 2007, 0 otherwise	.052		.053	
W15	1 wave 2008, 0 otherwise	.049		.051	
W16	1 wave 2009, 0 otherwise	.049		.053	
W17	1 wave 2010, 0 otherwise	.046		.0494	

W18	1 wave 2011, 0 otherwise	.050	.0557
W19	1 wave 2012, 0 otherwise	.048	.0561
W20	1 wave 2013, 0 otherwise	.043	.0518
Industry	**Industry where the individual works**		
Industry1	1 for other industries, 0 otherwise	0.035	.036
Industry2	1 for agriculture industry, 0 otherwise	0.015	.008
Industry3	1 for energy industry, 0 otherwise	0.017	.006
Industry4	1 for mining industry, 0 otherwise	0.007	.001
Industry5	1 for manufacturing industry, 0 otherwise	0.259	.131
Industry6	1 for construction industry, 0 otherwise	0.208	.045
Industry7	1 for trade industry, 0 otherwise	0.091	.163
Industry8	1 for transport industry, 0 otherwise	0.064	.034
Industry9	1 for financial industry, 0 otherwise	0.035	.047
Industry10	1 for services industry, 0 otherwise	0.271	.529

Appendix B - Figures

Figure B.1: Wage gaps in Europe, 2014



(¹) Enterprises employing 10 or more employees; NACE Rev. 2 B to S (-O).
(²) Provisional data; Ireland: 2012 data
(³) Estimated data
(⁴) 2013 data
No data for Greece

Notes:

i. Wage gaps in Europe 2014, measured as the difference in wages between male and female employees. Measured as average female gross hourly earnings as a percentage of male gross hourly earnings.
ii. **Source:** Eurostat (2016, March 1). Gender pay gap statistics. Retrieved from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics.