LUND UNIVERSITY
School of Economics and Management

Returns to schooling and potential signalling effects: Estimates based on ISSP data on Sweden

Carl-Gustav Schöön & Emil Bergman

Supervisor: Petra Thiemann

Assistant supervisor: Sara Moricz

Department of Economics, Lund University

NEKH03, Bachelor’s Thesis

May 2018

1. Abstract

This thesis estimates the returns to schooling in Sweden for working individuals by applying the Mincer equation on three datasets from the International Social Survey Programme and compare the results to assess potential trends. It also investigates if there is any difference in returns to schooling for men and women and if there seem to be any signalling effects present. The estimates indicate that the returns to schooling are around 4 percent in Sweden during the period of 1997 to 2015. Furthermore, our results suggest that the returns to schooling remains relatively constant during the entire period. The estimates do appear to differ between males and females, with the male estimates being higher for all years. The results also indicate that we might observe some signalling effects as the coefficients to the dummy variables for obtaining a University degree are statistically significant. However, this might be due to other factors. The resulting estimates are likely somewhat biased upward due to the innate ability of the individual not being accounted for and thus producing an ability bias.

Keywords: Returns to schooling; Signalling effects of education; Mincer equation; Expected wage; Ability bias; years of schooling

Content

2. Introduction

Have there been any change in the returns to schooling since the beginning of the millennia, and if that is the case, in which direction is the change? The purpose of this thesis is to investigate the development of returns to schooling in Sweden from 1997 until 2015. To accomplish this, we will apply the Mincer equation[1] (*Equation 1*) on three datasets from 1997, 2005 and 2015 and investigate whether there is a trend in returns to schooling, that is the private returns to one additional year of education. We will use the same criteria as Palme and Wright which is the years of schooling and the highest achieved level of education as explanatory variables, though estimations based only on years of schooling will be utilized initially. The thesis will assess how having a degree corresponds to the expected wage, as it is often proposed in signalling theory that the degree is of importance due to it signalling productivity. Hence, we will extend the basic Mincer equation (*Equation 1*) with dummy variables for the highest obtained educational degree to study whether it is indeed positively correlated with the earnings of the individual. Furthermore, we will look at whether the results are affected by the sex of the individual. Thus, the Mincer equation (*Equation* 3) with a dummy variable denoting the sex of the individual is applied on the datasets and two separate Mincer equations[2] are applied to the subsets of males and females.

To our knowledge, no estimates of the returns to schooling in Sweden utilizing data more recent than the mid-2000s exists. As returns to schooling concerns the average payoff to the individual it is a highly relevant topic for both individuals and policymakers evaluating the benefits of schooling. The Swedish estimates should be of interest as we can observe an increasing number of individuals acquiring higher levels of education (Statistics Sweden, 2017, p. 2). Hence estimates of the wage premium an individual has from an additional year of schooling is of utmost interest. Whether there is a trend in the returns to schooling should be of interest from a policy perspective as any policy aiming to encourage individuals to, for example acquire more schooling, would have to examine the trend to determine the appropriate measures to be taken. The significant coefficient for the University degree dummy variable also suggests some signalling effects might be present which could raise policy incentives to increase the ratio of students who graduate.

[1] The mincer equation is an OLS regression with the logarithmic estimated wage as a dependent variable and years of schooling, experience, experience squared and potential control variables as independent variables. The reference group in this equation represents the expected wage for a person with zero years of schooling and zero experience.

[2] These are based on equation 1 but run on separate subsets.

2.1 Previous literature/research

Several relevant studies have estimated the returns to schooling, both internationally and in Sweden, this section will assess the methods, results and commonly encountered problems in the previous research on the topic of returns to schooling. It will mainly focus on research concerning Sweden but will also to an extent assess research with a more international focus. These previous results, methods and previously encountered problems will be of importance for the thesis as they provide a comparison to our results, as well as establish methods, problems with certain approaches as well as overarching problems with estimations of returns to schooling. All of this will need to be assessed to do a satisfactory comparison of the results later.

There are a few recurring methods for estimating returns to schooling evident in these previous studies, including the popular Mincer equation, which is an OLS estimation, and Instrumental Variables estimations (Card, 2001, p.6). An important trend observed in previous research is that these IV estimates tend to be as high, and often as much as 20 percent higher than their corresponding OLS estimates (Card, 2001, p. 29). This is of importance since the Mincer equation, which is an OLS estimation, would thus likely yield lower estimates than an IV estimate of the returns to schooling on the corresponding variables.

A recurring problem in the previous literature on returns to schooling is the ability bias which affects the resulting estimates of an OLS regression and results in an upward bias in the estimates. Gary Becker constructed a model which describes this bias well. In this model the marginal benefits an individual's experience from acquiring more schooling increase in tandem with ability (Blackburn & Neumark, 1993, p. 5). The model also incorporates opportunity to reflect individual specific differences which make the marginal cost of schooling lower (Blackburn & Neumark, 1993, p. 5). Thus, the model predicts that if there does not exist a large enough negative correlation between opportunity and ability, workers with higher ability levels will tend to choose more schooling, hence an upward bias arises in the estimates of the returns to schooling (Blackburn & Neumark, 1993, p. 5). Thus, the ability bias is troublesome on a multitude of levels since the difference in ability will affect the return on education, but the magnitude of the ability bias problem might also differ as a response to innate ability (Nordin, 2005, p.2). Empirically the returns to schooling estimates appears to drop by 20 percent on average and by up to 25 percent at most, when a measure of individual ability was included in the regression (Kjellström, 1997, p. 30), thus further supporting the general consensus about the presence of an upward bias in the returns to schooling estimates. In a study from 2008 Nordin raises some criticism towards the standard returns to schooling estimates (Nordin, 2008, p.3). The criticism stems from the fact that the standard returns to schooling measures the average returns to schooling, while the actual returns probably vary depending on the innate ability of the individual (Nordin, 2008, p.3). As such the true returns could potentially be significantly larger for highly able individuals and significantly lower for low ability individuals (Nordin, 2008, p.3).

To accommodate for the aforementioned ability bias and thus attempt to examine a more causal relationship between earnings and schooling, later research often utilizes different supply side attributes of the educational system as Instrumental Variables to make an estimation of the returns to schooling (Card, 2001, P. 1), these studies usually yield higher values for the returns to schooling than OLS regressions do (Card, 2001, p. 1), thus slightly decreasing the comparability of the measurements (Card, 2001, p.29). They are also both place and time specific which to an extent limits their ability to explain the general situation.

Among the more recent studies of the returns to schooling concerning Sweden on this subject is a paper done by Mårten Palme and Robert E. Wright published in 1998, in which they estimate change in returns to schooling in Sweden with measurement points from 1968, 1981 and 1991 (p.1). They do this by using the Mincer equation and a cubic spline function (Palme & Wright, 1998, p.4), both producing similar results (Palme & Wright, 1998, p.10). They exclusively use data from the *Swedish Level of Living Surveys* which is carried out by the Swedish Institute for Social Research at Stockholm University (Palme & Wright, 1998, p. 3). In their study of returns to schooling in Sweden for the years 1968, 1981 and 1991 Palme and Wright estimated returns to schooling for men to be 8.2 percent, 4.2 percent and 4.0 percent for each year respectively and 7.4, 3.5 and 3.5 for each respective year for women (1998, p.6). These estimates were produced using the same Mincer equation (*Equation 1*) as we will be using to produce our estimates and thus provide a reasonable point of reference for our results. The estimates by Nordin are in the range of 3 percent up to about 8 percent depending on score on the military battery test and based on data from 2001 (Nordin, 2008, p.13) while the estimates obtained by Björklund et al. range from to approximately 4.5 percent in 1991 to about 6 percent in their last measurement year, 2000 (Björklund et al., 2004, p.44).

Palme and Wright's study concluded that that the return to schooling in Sweden appears to have decreased drastically between the measurement years, 1968, 1981 and 1991 (Palme & Wright, 1998, p.6). This change is observable when allowing for signalling effects as well (Palme & Wright, 1998, p.4). However, Björklund et al. in their 2004 report for the Institute for Labour Market Policy Evaluation finds that the returns to schooling has risen during the second half of the 90s and end up at about 6 percent in 2000 (2004, p. 45.) Furthermore, Nordin finds that the returns to schooling in Sweden have in fact increased after 2000s as well (2008, p.8). In regard to this development, it is interesting to note the relatively low return to higher education in Sweden compared to many other western countries (Björklund et al. 2004, p.47). Furthermore, the change in returns to schooling over the period 1968 to 1991 is not evenly distributed over the different schooling levels (Palme & Wright, 1998, p.5), the largest relative decline being seen in the return to university studies (Palme & Wright, 1998, p. 4), meaning a linear return to schooling estimate is somewhat misleading.

The returns to schooling estimates in Björklund et al.'s study is based on standard Mincer estimates which produce a linear years of schooling estimate (Björklund et al. 2004, p. 44). This highlights a limitation in the standard Mincer equation (*Equation 1*), that it only produces linear estimates of returns to schooling suggesting each year corresponds to the same effect on the expected wage, which Björklund and Kjellström shows does not empirically seem to hold true (2002, p. 15). Thus, it is yielding a very general estimate of the effect of schooling.

Previous research has concluded that the ability bias is the largest problem in estimations of returns to schooling, as it decreases the explanatory power of the models utilized to estimate the returns to schooling as well as bias the resulting estimates making them less reliable. This error margin is important to account for when assessing estimates of returns to schooling. Further both Nordin (2008, p.8) and Björklund et. al. (2004, p.45) has found positive trends in returns to schooling around the year 2000, while Palme and Wright found a decreasing trend from 1968 to 1991 (1998, p.4). These results are important points of comparison for our own results. Further this section has concluded the prevalence of the Mincer equation to estimate returns to schooling which makes it a reasonable tool to use as it makes for more comparable results.

3. Theory and model

To maintain comparability with the results from previous research this thesis will utilize the Mincer equation. The Mincer equation is an OLS regression with the logarithmic wage (*Lnwage*) as the dependant variable and *Schooling*, Experience (*Exp)* and Experience squared (*Exp2*) as independent variables. To properly interpret the resulting coefficients, it is important to understand the theoretical framework the model itself is based upon which will be elaborated in this section. The results in this thesis will also be assessed using human capital theory and signalling theory. Thus, these will also be discussed in this following section.

The thesis will be based upon estimates of the returns to schooling constructed through use of the Mincer Equation. As noted by Jacob Mincer (1974) there is a positive, although not entirely simple, relationship between the education of an individual and the individuals later earnings (p.1). This may be understood as a result of a productivity enhancing effect of schooling (Mincer, 1974, p. 1). In accordance with human capital theory, that states that schooling enhances the productivity of the individual in a direct manner and that the wage is supposed to mirror the productivity of the worker (Rospigliosi, 2014, p.3). However, this relationship does not exist in a vacuum. Thus, there is a fair amount of noise in the relationship between years of schooling and later earnings, due to factors such as, varying quality of education, productivity enhancing experiences acquired elsewhere and many other factors. This result in a relatively weak observed correlation between years of schooling and earnings (Mincer, 1974, p. 1).

The Mincer equation is based on a rather early form of the Human capital model which is then expanded upon to handle earnings differentials amongst different cohorts, in the model accomplished via the experience variable (Mincer, 1974, p.1). The model defined by Mincer, states decisions on investments in schooling are based on the present value of future earnings (Mincer, 1974, p.8). The essential assumption that is made in the theoretical aspects of Mincer's model is that investments in human capital have a positive influence on earnings through increased productivity but also consumes time and thus in a sense have some degree of opportunity cost (Mincer, 1974, p.12). The use of the experience variable helps incorporate the observed reality that individuals do invest in human capital even after finishing school, but that this occurs at a decreasing rate over the course of the individuals lifecycle (Mincer, 1974, p.14), to account for this nonlinearity the Experience squared variable is added (Mincer, 1974, p.84).

Mincer further elaborates on this, claiming part of the weakness in the correlation is to be attributed to the fact that years of schooling as a measurement does not incorporate a satisfactory measure of all the costs and quality aspects inherent in education (Mincer, 1974, p. 133). Another problematic aspect influencing the strength of the correlation is the occurrence of training or educational investments done by the individual after leaving school. Assuming that these post-schooling investments are not evenly distributed over the population, and that these investments are important for the wage level, the correlation between these investments and schooling does not appear to be strong. The result is a weakening of the correlation between schooling and earnings of an individual will continue to decay as experience grows. This is one of the reasons for the inclusion of the experience squared term (Mincer, 1974, p. 133).

The thesis will also analyse to what extent the fact that an individual has a degree affects the returns to schooling. The most common way to describe such a potential effect is via signalling theory. In this theory, when applied to the job market, if a worker applies for a job there is a great deal of uncertainty involved (Spence, 1973, p.2). This uncertainty arises as the employer lacks information about the productivity of the worker (Spence, 1973, p.2). Furthermore, the worker will likely not reveal the extent of their productive capacity during the initial time on the job either due to, among many factors, job specific skills that must be learnt which takes some time (Spence, 1973, p.2). Thus, the employer makes the decision on whether to hire the individual based on observable traits and features of the applicant (Spence, 1973, p.3). Having a degree is one such factor and therefor having a degree should, in theory, grant a significant advantage as it signals productivity due to the assumption that the cost of acquiring the signal, the signalling cost, would be higher for individuals with lower productivity (Spence, 1973, p.4).

We suspected that this signalling effect of the degree variables might vary depending on the educational level attained. To accommodate for these effects and study the correlation we will add dummy variables corresponding with the schooling levels obtained from the degree variable. The age composition of the groups composing the variables differ slightly as the variables are aggregated from degrees obtained during different configurations of the Swedish schooling system[3]. Currently the compulsory schooling is 10 years long (Palme & Wright, 1998, p.4) thus making the compulsory schooling correspond roughly to the age of 16. Some variations within the group likely occurs due to the aforementioned changes in the compositions of the schooling system. The first dummy category corresponds with vocational education, the education has changed slightly over the years, but has consistently been two to three years (GESIS Data Archive for the Social Sciences, 2015, p.6), the current version "Yrkesinriktat gymnasieprogram" (GESIS Data Archive for the Social Sciences, 2015, p.6), is a programme within the Swedish gymnasium (Skolverket, n.d.) and would thus correspond roughly to the age of 19. The second dummy variable is aggregated based on the different configurations of gymnasium education or equivalent. Currently gymnasium amounts to three years (Skolverket, n.d.), while the third dummy is constructed of the university degree observations. As university degrees vary in length and can be initiated at different ages there is no easy way to put a corresponding age on the *D_uni* variable. But as have been established, the gymnasium corresponds roughly to the age of 19 and individuals who are eligible can start university straight after graduating the Gymnasium. Considering that a bachelor's degree is three years (Uppsala Universitet, n.d.) the degree could be considered to correspond with the age of 22, however there is great variation in this and according to the OECD the average age of Bachelor graduates is 28 in Sweden (OECD, 2017 p.75).

When constructing *Equation 2*, we have chosen to use the different degrees corresponding with the compulsory education in Sweden, as the reference group. The reference group also includes the degree "folk high school"[4], this choice requires some further explanation as folk high school is not part of the compulsory education, but rather a voluntary education usually obtained at some time after the gymnasium, which is the Swedish equivalent of high school. The inclusion of folk high school into this category was motivated as the courses offered there are mostly focused on more artistic subjects and generally are not meant to train the student for a profession. The amount of observations affected by this choice is small though and should not have any significant effect on the result.

[3] For further information see section 4.1 describing the sorting and construction of the different educational level variables.

[4] The translation into English done in the dataset is somewhat misleading as it is a literal translation of the Swedish word Folkhögskola, which is not in any way associated to the more international term high school, it is merely a non-compulsory school offering courses in different fields, usually within the artistic subjects.

To be able to assess any effects being male or female might have on the estimated returns to schooling the theoretical framework must be established. Signalling theory suggests that there are two types of observed attributes to the individual. These are signals which are alterable by the individual, and indices, for example sex, which are unalterable (Spence, 1973, p.3). The implications of this is that, given that the probability of finding an equally productive individual in a random sample among males and females is the same, the sex of the individual will not give the employer any information about productivity (Spence, 1973, p.15). Thus, it should have no effect on the wage or employment (Spence, 1973, p.15). However, there are externalities present as the individuals signalling decisions have an impact on the market data for the group the individual belongs to (Spence, 1973, p.16). Thus, if the employer, under incomplete information, makes a decision conditioned on education and sex, the differences in educational attainments among males and females will act as a signal for that group and the individual will be assessed based on the average member of the group conditioned on sex and education (Spence, 1973, p.16).

In this section the Mincer equation along with human capital theory and signalling theory have been assessed and elaborated on. The Mincer equation itself is based on the assumptions made in human capital theory but allows for investigations of potential signalling effects via the addition of dummy variables. Part of the strength of the Mincer equation is the versatility of it. It does however have a relatively low explanatory power manifested by low R-squared values. Thus, it is important to note that there is a rather large amount of noise in the model, a lot of it stemming from other productivity enhancing traits of the individual in accordance with human capital theory, alternatively it is feasible that the individual has more traits signalling productivity other than schooling. To properly assess the resulting coefficients from the Mincer equation the unexplained variance should be noted.

4. Data + descriptive statistics

The data in this essay is comprised of three data sets from the International Social Survey Programme, ISSP. These are Work Orientations II, from 1997, Work Orientations III, 2005, and Work Orientations IV from 2015. (Umeå Universitet, n.d.) Work Orientations I have been excluded as it does not include data for Sweden (Umeå Universitet, n.d.). The data sets are based upon survey data collected by organizations in each country respectively (Umeå Universitet, n.d.). The data for Sweden was collected from surveys performed by Statistics Sweden for Work Orientations II, and SIFO for Work Orientations III and IV (Umeå

Universitet, n.d.). The data processing and sorting as well as the regressions were conducted in R. The data collection and how the sorting into our final subsets of them is highly important for the thesis as it gives the reader a transparent view of what we are examining. Hence this section will explain the sorting process, which observations where selected for the subsets we apply the Mincer equation to and why these where selected.

The International Social Survey Programme, ISSP, is a collaborative project with participants from most parts of the world, although we will here exclusively use their data related to Sweden. The aim of the project is to create internationally comparable survey data on individual attitudes. As Sweden did not join the project until 1992 our data has been from the related studies conducted after its entry. Since the study done in 1997 and those conducted in 2005 and 2015 where conducted by different organisations, the variable names change slightly between the data sets, this is also affected by a revision of the guidelines for how the variables are documented, done in between the studies. This has resulted in the background information for the two later studies being much more extensive than for the data set created in 1997[5] (Umeå Universitet, n.d.).

It should be noted that since the datasets utilized are based upon survey data some measurement error is to be expected assuming that not all individuals have given correct answers. This might be further reinforced by the way many of the questions, regarding for example work hours have been asked, that is, asking for the average amount of work hours during a week. Furthermore, there are evidence that the reliability of the self-reported values regarding years of schooling, generally tend to be somewhere in the span of 85-90 percent (Card, 2001, p.9). While these results do not specifically concern our datasets, they do suggest that there is some error margin present.

We also observe some change in the age compositions of the individuals participating in the survey between each year. The main observed change in this regard is the increase in people over the age of 65 in the data set from 2015 (Appendix 3: Figure 3.1). This will likely affect our results somewhat as these individuals are likely to be retired and thus work less or not at all. Furthermore, it should be noted that some changes to the structure of the educational system in Sweden took place over the last half century, resulting in many different answers

[5] For further information on the background variables see Appendix 1: Variable list.

under the *degree*-variable, these are roughly equivalent, but it is not unlikely that some margin of error in the returns to schooling estimates is attributable to variations of the educational structure and number of years required to obtain an equivalent degree.

In sorting the datasets, we choose to exclude observations reporting 0 work hours. The arguments for excluding observations reporting 0 workhours on an average work week is mainly stemming from the aim of the thesis. The purpose of this study is to examine the supposed positive correlation between schooling and estimated wage, since we are specifically interested in the correlation with wage and not income, it appears feasible to exclude individuals reporting that they do not work, as they then do not have a wage. However, this does not appear to be the standard way to restrict the datasets while conducting this type of Mincer equation. Palme and Wright restrict their data by an age interval defining the usual age of the labour force (Palme & Wright, 1998, p. 3), this method is also utilized by Björklund and Kjellström (2002, p. 3), thus excluding individuals who are likely not in the labour force and including observations of unemployed individuals within that age span. We argue that this adds more noise than necessary in our case as our datasets contain a measure of average weekly workhours and thus provide a more reliable instrument for who is in the labour force and who is not, instead of using the age span as a proxy. Nordin uses a slightly different restriction using three restricting criteria based on income, age as well as whether the individual has been studying during the last year (Nordin, 2008, p.3), thus suggesting an overall aim to study only individuals with a wage.

Observations reporting 0 work hours on an average week, were excluded as only the working individuals is of interest in this thesis. After excluding these observations, the aggregate data set were relatively unaffected by missing values, with 90.565 (*Appendix 4: Figure 1.1*) percent of the observations having no missing values in any of the variables accounted for in the regression nor in the variables that these are based upon. One can also observe a strong correlation between the missing values. Observations missing values in *Schooling*, also missing in *Exp* and *Exp2*, the two experience variables (*Appendix 4: Figure 1.1-1.4*). This result is to be expected as the experience variables are constructed via the *Schooling* variable (*Appendix 1: Variable list*). As the *Whour* (hourly wage) and *lnwage* variables are both constructed based upon the *Wage* variable the same holds true for these (*Appendix 1: Variable list*). Furthermore, some missing values are observed independently of the other variables in *Degree* and *Wrkst* (work status variable) these however account for such a low percentage of the overall result, 0.771 percent and 0.881 percent respectively, that they should not have any substantial effect on the result (*Appendix 4: Figure 1.1*).

We started with a little more than 3800 Swedish observations in total. These were then quite evenly distributed between our sets with 1275 observations from 1997, 1371 from 2005 and 1162 from 2015 (*ISSP Research Group, 1999 & ISSP Research Group, 2013 & ISSP Research Group, 2017*). When correcting for missing values and removing observations who reported 0 work hours we received the following samples: 1997 had 874 observations, 68.5 percent, of the Swedish observations left after adjusting for 0 work hours (*Appendix 3: Table 1.5),* 2005 had 722 observations, 0.526 percent of the observations left (*Appendix 3: Table 1.5*) and the 2015 had 684 observations accounting for 58.8 percent of the observations left (*Appendix 3: Table 1.5*). The Mincer equations are then applied to these observations. This massive loss of observations was to be expected since the datasets contained people younger than 20 and older than 65, that is, people who are likely to be out of the labour force. A part of the loss of observations is because the sorting also removed the unemployed observations. Another cause for this loss was also due to the fact that our sets contained some missing values. Despite the large decrease in the number of observations in the data, the datasets are rather reliable with a sufficient number of observations, and low percentage of observations affected by missing values, as indicated by *Appendix 4: Figure 1.1*[6]. The variables most affected by missing values are *WHours* and *Lnwage*, not surprising as they, by construction will correlate completely as *Lnwage* is constructed from the variable *WHours*[7] *(Appendix 1: Variable list)*. Naturally the same goes for the variables *Schooling* and *Exp* and *Exp2* as the experience variables are constructed using the values from *Schooling*[8] *(Appendix 1: Variable list)*. It should be noted that there is some variation in the portion of observations containing missing values across the different data sets. While the total data set has around 90

---

[6] *Appendix 4: Figures 1.1-1.4* describes the portion of missing values in our data sets and how missing values are distributed across our variables. The bar chart shows how large percentage of each variable is affected by missing values. The blue and yellow box on the right shows all possible combinations of missing values and variables. To see the effect of missing values, the box is to be read as rows, the completely blue row is the case where there is no missing values, and the values at the far right show how large a portion of the dataset this configuration accounts for. Each column represents a variable and each yellow cell tells that this instance is affected by missing values so if read as rows each yellow cell indicates that this row shows a configuration where that variable has missing values.

[7] As visible in the bar chart for *Appendix 4: figure 1.1* they correlate completely and when studying the distribution, one can observe that there are no configurations in the total data set where there is a missing value in *Whours* or *lnwage* alone.

[8] Once more, looking at the distribution chart one can conclude that there are no configurations where any of these three variables are the only variable having missing values.

percent observations without missing values (*Appendix 4: Figure 1.1*)[9], the percentage decrease to around 71 percent for the 1997 data set (*Appendix 4: Figure 1.2)*[10], just under 67 percent in the 2005 data set (*Appendix 4: Figure 1.3*)[11]. The portion of observations without missing in the 2015 data set is about 82 percent (*Appendix 4: Figure 1.4*)[12]. The aforementioned change in the portion of missing values mainly stems from changes in the amount of missing values in the variables *WHours* and *lnwage*, which decline drastically in the 2015 dataset as compared to the 2005 and 1997 data sets[13]. The higher portion of observations unaffected by missing values in the 2015 is likely due to the increasing frequency of the answer 0 to weekly workhours in that dataset, as evident by *Appendix 4: Figure 2.1.* Together with the high ratio of unaffected observations in the 2015 dataset it appears likely that part of the explanation for the variation in the ratio of observations unaffected by missing values over the datasets are due to how much of the unemployment is captured as 0 workhours and how much is coded as NA due to respondents not answering the question.

The descriptive statistics in *Tables* 1.1-1.4 provide a good overview of the differences between the datasets. These tables are based upon the datasets that have had non-working observations removed, i.e. the same sets that we will use in the statistical test.
Initially we can conclude that the distribution between the sexes are even, it ranges from 49 percent women (*Appendix 3: Table 1.2*) to 53 percent women (*Appendix 3: Table 1.3*) depending on the dataset. The age distribution and the workhours distributions are also similar across the datasets. The mean age ranges from the lower 43.62 in the 1997 dataset (*Appendix 3: Table 1.1*), to the higher 46.70 in the 2015 dataset (*Appendix 3: Table 1.3*). In the same manner the mean work hours are consistently around 38 hours per week with the low end being 37.64 in the 2005 dataset (*Appendix 3: Table 1.2*) and 38.64 for the 2015 dataset (*Appendix 3: Table 1.3*).

---

[9] As visible in the value beside the entirely blue row in *Appendix 4: figure 1.1*, representing the configuration without missing values.

[10] Given by the value next to the entirely blue row representing the configuration where there are no missing values in any variable in *Appendix 4: figure 1.2*.

[11] Given by the value next to the row showing the configuration where there are no missing values in any variable in *Appendix 4: figure 1.3*.

[12] Given by the value next to the row showing the configuration where there are no missing values in any variable in *Appendix 4: figure 1.4*.

[13] This conclusion is apparent if one compares the scale of the bar chart and the entry for *lnwage* and *WHours* between *Appendix 4: figures 2, 3 and 4*.

The 2015 dataset appears to differ compared to the other two datasets. Especially with respect to *schooling*, *wage* and the distribution of level of schooling obtained. While both 1997 and 2005 have very similar mean for *Schooling* and degree distribution; the *schooling* mean is 12.16 and 12.49 years of schooling respectively for each dataset (*Appendix 3: Table 1.1-1.2*). The highest obtained level of education is quite evenly distributed for both the 1997 and the 2005 set, approximately 25 percent on each (*Appendix 3: Table 1.1-1.2*). The interesting differences between these two datasets and the 2015 dataset is the increase in the amount of highest degree of education obtained, i.e. higher means for *Schooling* and *D_Uni (Appendix 3: Table 1.1-1.3)*. The 2015 mean for *schooling* is 14.19 years and the share of people with university degree have increased to almost 50 percent, meanwhile the share with compulsory school have decreased to 7 percent (*Appendix 3: Table 1.3*). The last notable difference is the drastic increase in *Wage*, it increases from a mean that is little more than 21 000-23 000 SEK to almost 36 000 SEK (*Appendix 3: Table 1.1-1.3*).

These changes are to an extent reasonable. The level of education has increased steadily in Sweden during the last years, this increase account for both the years of schooling and the highest level of education obtained (Statistics Sweden, 2017, p. 2). The drastic increase in the mean wage is probably to some extent correlated with this increase in education, in combination with a nominal and real increase in wages (Carlgren, 2017). But it is likely that some part of it is due to measurement errors.

Another area in which the datasets differ is the reported number of average work hours per week, once more, it is the 2015 set that is unusual (*Appendix 4: Figure 2.1*). It has a much larger proportion of the observations that report zero hours now compared to 1997 and 2005 (*Appendix 4: Figure 2.1*). However, this is quite understandable if one looks at the data. When inspecting the figures mapping the missing values, one can observe a drastic decrease in the percentage of observations containing missing values in the *Hours* variable for the 2015 set (*Appendix 4: Figure 1.2-1.3*)[14]. Furthermore, the ratio of observations with 0 work hours is larger for the 2015 dataset compared to the other two data sets (*Appendix 3: Table 1.5*)[15].

[14] Visible in the bar chart as the 1997 chart shows approximately 18 percent of observations had missing values in the hours variable (*Appendix 4: Figure 1.2*), the bar chart for 2005 has a corresponding value of approximately 25 percent (*Appendix 4: Figure 1.3*) while the corresponding value for the 2015 dataset is approximately 6 percent (*Appendix 4: Figure 1.4*).

[15] Based on own calculations on the data set (all observations – observations remaining after removing 0 work hours) / all observations. Based on this the ratio of observations with 0 workhours for 1997 is approximately 23.8 percent, the corresponding values for 2005 and 2015 is 25.0 percent and 37.7 percent respectively.

Thus, it could be the case that some of the loss to 0 workhours in 2015 has captured some of the loss to missing values in the other datasets.

4.1 Sorting

When sorting the data into the four educational level variables utilized in the regressions to study potential signalling effects we followed the general pattern made by Palme and Wright in their 1998 study on the returns to schooling in Sweden (Palme & Wright, 1998, p.1). This division appears to be consensus in this field. The categories used in their study are exhibited in the quotation below:

> *"SI: Basic compulsory (e.g. folkskola, junior high school, realexamen, grundskola, högre folkskola, flickskola, folkhogskola), which is currently ten years of schooling; (2) S2: Vocational education for at least 1 year in addition to the basic compulsory education level SI: (3) S3: Completed gymnasium (c. high school) an/or Vocational education in addition to high school. Gymnasium is three years' schooling beyond basic compulsory education and is required for university entry; and (4) S4: Completed university education." (Palme &Wright, 1998, p.4)*

To accommodate this categorisation of our variables we utilized the translation of these provided from the ISSP background variable documentation (GESIS Data Archive for the Social Sciences, 2005, p. 6). Thus, we end up with categories S1: Primary or comprehensive school, Folk high school, Alternative secondary school, Lower secondary school, Incomplete primary or comprehensive school, primary/comprehensive school pre 1962 (6-8) years, primary/comprehensive school post 1962 (9 years). S2: Vocational school (post 1972), Vocational school (pre 1972), Vocational school (1972-1992), Vocational school (post 1992), Vocational school, Vocational school 1963-70, Vocational school post 1992. S3: 3 or 4-year gymnasium (academic track), gymnasium (academic track post 1992), gymnasium, academic track post 1992, higher secondary school. S4: University degree, 3 years or more, doctor's degree, university degree, less than 3 years, university degree. The sorting is summarized in table 1.6 below.

Table 1.6 Levels of education

S1	S2	S3	S4
Primary or comprehensive school	Vocational school (post 1972)	3 or4-year gymnasium (academic track)	University degree, 3 years or more
Folk high school	Vocational school (pre 1972)	gymnasium (academic track post 1992)	doctor's degree
Alternative secondary school	Vocational school (1972-1992)	higher secondary school	university degree, less than 3 years
Lower secondary school	Vocational school (post 1992)		university degree
Incomplete primary or comprehensive school	Vocational school		
Primary/comprehensive school pre 1962 (6-8) years	Vocational school 1963-70		
primary/comprehensive school post 1962 (9 years)			

These correspond well with those utilized by Palme and Wright and should thus yield comparable results.

The data comprised from the three datasets from the International Social Survey Programme (ISSP) where sorted into subsets containing all Swedish observations, as the focus of this thesis is Sweden, and from there observations reporting to not work where excluded. There is some variation in demographics, most notable within the mean age, wage and amount of schooling between the datasets. These differences are important as they can affect the result and how it could be interpreted due to each coefficient being derived from the datasets due to the observations providing a form of sample selection. The missing values and where they appear are an important aspect as well as too large a proportion of missing values within a single variable would be highly problematic for drawing any kind of conclusion.

5. Method

As stated in the theory and model section the thesis results are based on estimates of returns to schooling derived through the Mincer equation, this equation will be elaborated on in the following section. Furthermore, econometric validation for using the model on the datasets in question are provided due to the importance of validating that the results are not econometrically unreasonable and that they are not overtly biased due to the sample in the datasets. The standard Mincer equation (*Equation 1*) and the expanded versions (*Equation 2 and Equation* 3) of it utilized in this thesis use some variables not in the original dataset, these had to be created using variables in the datasets. Hence it is important that these are thoroughly explained in order for them to be interpretable.

5.1 Mincer equation

The Mincer equation is constructed as follows:

(*Equation 1*)

$$Ln(wage) = \beta_0 + \beta_1 S_i + \beta_2 Exp_i + \beta_3 Exp_i^2 + \varepsilon_i$$

Where *ln(wage)* is the logarithm of earnings (*Appendix 1: Variable list*), S_i is years of schooling *(Appendix 1: Variable list*) and Exp_i is years of potential experience (*Appendix 1: Variable list*). The β_0 is the intercept in the equation and is to be interpreted as the logarithmic wage of a person with zero years of schooling and with no experience. β_1, β_2 and β_3 measure then the estimated return to schooling and experience. The Exp_i^2 variable (*Appendix 1: Variable list*) is included, as it both, improve the explanatory power of the model, but also as stated by Mincer, the expected payoff to experience is likely not a linear function and thus the addition of the Exp_i^2 variable helps depict work experience more accurately (Mincer, 1974, p.84). The years of schooling variable is very straightforward since we already had that information in our data. Experience was not a part of the original dataset, so we defined it as:

$$Exp = Age - Schooling - 6$$

Mincer defined the experience variable by subtracting the age of completion of schooling from age (Mincer, 1974, p.47). The minor difference is that Mincer had access to information on the completion of schooling, but by subtracting 6 and years of schooling from the age a variable that is sufficiently equivalent is obtained, since children in Sweden normally start school when they are 7 seven years old (Skolverket, 2015). This proxy for work experience assumes that the individual obtained no applicable experience before beginning of schooling and did not work during schooling. The first assumption should be indisputable, however the second one is perhaps less plausible, but it should not affect our estimations to a large extent. This estimation method generated a couple of negative experience observations, these are recoded to zero. This should not have any effect on the results since it was only 2 to 5 cases in our datasets that was affected by this.

We also did an extended Mincer with dummies for different levels of obtained education, resulting in the following model:

(*Equation 2*)

$$Ln(wage) = \beta_0 + \beta_1 S_i + \beta_2 Exp_i + \beta_3 Exp_i^2 + \beta_4 D\_Vocat_i + \beta_5 D\_Gym_i + \beta_6 D\_Uni_i + \varepsilon_i$$

D_Vocat is the dummy representing the vocational training in addition to the reference group (compulsory education), *D_Gym* is the dummy for completed gymnasium and *D_Uni* is then the dummy for a University degree (*Appendix 1: Variable list*). Thus is β_4, β_5 and β_6 the estimated return to schooling depending on the highest obtained education.

In excess to these two we also did an extended equation using sex as a dummy:

(*Equation 3*)

$$Ln(wage) = \beta_0 + \beta_1 S_i + \beta_2 Exp_i + \beta_3 Exp_i^2 + \beta_4 Female_i + \varepsilon_i$$

Here the dummy, *Female (Appendix 1: Variable list)*, is meant to capture differences based on sex in the expected wage and β_4 is the estimated difference. In this model the reference group is men. In addition to this we also redid the ordinary Mincer on subsets consisting of men and women separately. These two will give us approximations on the difference between men and women regarding both the estimated return to Schooling and the intercepts, i.e. the expected differences in logarithmic wage when schooling and experience is the same. This will give us a helpful estimation on the unexplained differences between sexes.

5.2 Variables

The main variables used in the regression are defined in detail in appendix 1: Variable list, since the survey changed slightly over the course of the three measurement points each variable will be defined for each measurement point. Thus, both the original variables and the created variables are defined, to give a transparent picture of the data underlying this thesis.

There are a few ways to handle the upward bias caused by omitted ability variables. The most recurring ones being to add proxies for ability, comparing differences in differences in relation to a policy change or adding variables related to ability to capture some of the effect. We do not have a satisfactory way of correcting for ability bias due to restrictions in the data material provided. There are some arguments for including variables relating to family background into the regression as it in a sense may act as a proxy for the individual's ability, however the data provided in our data set merely provides information on the parents country of origin, and the self-reported societal status class, both of these are unreliable approximations of the individuals ability as they do not offer any satisfactory indicators on the genetic or environmental circumstances concerning the individual. Furthermore, it has been shown that adding a family background variable does little in the way of adjusting for the upward ability bias and does not have any considerable impact upon the returns to schooling estimates (Kjellström, 1997, p. 19). Due to this we have chosen to opt for a regression model without a family background variable.

5.3 Econometric validation

The R^2-score for our models are low, ranging from 0.169 for the total data set (*Appendix 3: Table 2.4*), 0.1892 for the 1997 data set (*Appendix 3: Table 2.1*), 0.1521 for the 2005 (*Appendix 3: Table 2.2*) and the low 0.0949 for the 2015 data set (*Appendix 3: Table 2.3*), meaning that our variables tend to describe the variation in *lnwage* to a rather limited extent. However, viewed in comparison to older research, this is to be expected since wage is dependent on a multitude of various factors other than schooling. While the actual productivity enhancing capability of schooling will vary greatly as well, due to quality differences in the education (Mincer, 1974, p.1). There is also the problem with measuring ability, as human capital theory states wages are set according to the individuals productivity (Rospigliosi, 2014, p.3), and previous research shows that including a measure of ability can potentially decrease the returns to schooling estimates by as much as 25 percent (Kjellström, 1997, p.30), thus likely attributing to the low R^2 value when excluded as it is then likely to explain a large proportion of the wage. In general, the R^2 values are mostly in line with that in previous research, which falls in the range of 0.189 to 0.269 in Nordin's study (2008, p.7). Kjellström's R^2 values are also 0.177 on the low end and tend to be about 0.20 (Kjellström, 1997, p.20). When running a regression on the aggregated dataset including the dummy variables *Sex, D_Vocat, D_Gym, D_Uni, D_mar, D_sep, D_wid, D_sub, D_urb, D_smagot, D_south, D_west, D_northmid, D_north, D_stock, (Appendix 1: Variable list)* in addition to the varaibles to the standard Mincer equation (*Equation* 1) the R^2 value increased to 0.2532 (*Appendix 3: Table 2.5*).

However, there is an additional way to control our models utilizing the residuals we create in each model. These tests will allow us check if the econometric assumptions are met and see whether there is anything problematic with our data and model. In R we use a built-in diagnostic test that generates four diagnostic plots; Residuals vs Fitted, Normal Q-Q, Scale - Location and Residuals vs leverage. We ran these tests on all our regressions resulting in the diagnostic plots which can be seen in *Appendix 2: Econometric validation*.

The purpose of the first test, Residuals vs Fitted, is to provide us with proof that the residuals does not have a non-linear pattern. Scale-Location is for determining if our models are affected by heteroscedasticity, i.e. if our assumption of equal variance holds. The normal Q-Q plot shows if the residuals are normally distributed. The Residuals vs Leverage helps us reveal any influential cases, i.e. observation that deviates from the trend to the extent that it strongly influences the result. Overall the plots seem to give validation to our models by confirming that the assumptions for linear regression are met.

The residuals in the Residuals vs Fitted is supposed to be distributed around a horizontal line without any sort of pattern. The Scale-Location plot shows whether the regressions fulfils the assumption of equal variance, the ideal case is for the residuals to be equally spread around the x-axis. If the residuals follow a straight line in the Q-Q plot that is a good sign that the residuals are normally distributed. The last plot, Residuals vs leverage, helps detecting influential observations. If any of our observations are located past the dashed red curve the removal of that specific observation would result in an altered regression which is problematic.

Overall our regressions performed well *(Appendix 2: Econometric validation)*. There are a few exceptions though, some of the Residuals vs leverage seems to indicate that there are cases in which certain observations influence the regressions estimates *(Appendix 2: Econometric validation)*. The normal Q-Q plots show some deviation in the “tails” but this is to be expected and in our case the deviation is no cause for concern *(Appendix 2: Econometric validation)*. The Residuals vs fitted confirms that there is a linear relationship between predictor variables and an outcome variable *(Appendix 2: Econometric validation)*. Scale-location confirms that there is no problem with heteroskedasticity in our models[16] *(Appendix 2: Econometric validation).*

These tests conclude the section and suggest that the Mincer equation is applicable on our data, that the data does not contain any outlier that bias the data and that the resulting estimates should be valid. As the equations have been explained, the variables examined, and the created variables explained[17], the reader should have the necessary background information to assess the results following in the next section. However, while assessing the estimated coefficients the reader should note the probable presence of some extent of an upward ability bias as well as that the dependent variable *lnwage* is not the wage but the logarithmic wage.

6. Results

The necessary background information, in the form of previous literature, specifications of the equations and variables utilized as well as theoretical framework have been established. Hence the following section will describe and examine the resulting estimates of the returns to schooling resulting from applying the Mincer equation and its derivatives on each dataset respectively. These results will be the topic of the discussion section below where their implications will be discussed in more detail.

Our results from the Mincer equations support the theory that years of schooling is positively correlated with the expected logarithmic wage, the same holds true for experience.
A recurring problem with our ordinary Mincer models are the low R^2-values, the highest being approximately 0.19 (*Appendix 3: Table 2.1*), meaning that the model only explains 19 percent of the variance in the estimated wage. However, considering the amount of noise to be expected in a correlation of this kind, as Mincer noted, the correlation is bound to be weak (1974, p.1). If we view the R^2-values compared to those obtained in previous research by Nordin we notice that the R^2-values of our model falls in the same span as the models used by Nordin (2007, p.7), albeit in the lower end of the spectrum. Thus, the low R^2-values appears reasonable in the given context and should thus pose no problem.

[16] The only one that is troubling is the Mincer with dummy for sex that is performed on the aggregated dataset but this should not be of any concern for our results.
[17] The interested reader can assess the definition of all variables in Appendix 1: Variable list.

6.1 Returns to schooling estimates, basic Mincer

Table 1.7 Mincer on 1997, 2005, 2015 and aggregate dataset

	Equation 1			
	n	RTS	Std. Error	R-square
1997	874	0.04268***	0.005008	0.1892
2005	947	0.04077***	0.004496	0.1521
2015	684	0.03968***	0.005999	0.0949
Total	2518	0.04808***	0.002272	0.1690

The returns to schooling estimate obtained from the standard Mincer equation (*Equation 1*) for the 1997 dataset, corrected for zero work hours observations is 0.04268, about 4.27 percent per additional year of schooling (*Appendix 3: Table 2.1*), with experience having a very similar coefficient of 0.04141 suggesting a positive correlation of 4.1 percent increase in estimated wage per year of experience (*Appendix 3: Table 2.1*). The R^2-value is 0.1892 (*Appendix 3: Table 2.1*).

Applying *Equation 1* on the 2005 dataset results in returns to schooling estimates of 0.04077 (*Appendix 3: Table 2.2*), just above 4 percent while the corresponding estimate for experience is 0.02835 (*Appendix 3: Table 2.2*), approximately 2.8percent (*Appendix 3: Table 2.2*). The R^2-value is 0.1521 (*Appendix 3: Table 2.2*) thus explaining around 15.2 percent of the wage (*Appendix 3: Table 2.2).*

Corresponding estimates for the 2015 dataset is 0.03968 (*Appendix 3: Table 2.3*) for schooling, about 3.97 percent, 0.02472 (*Appendix 3: Table 2.3*) for experience. The R^2-value is relatively low compared to the other regressions, 0.0949 (*Appendix 3: Table 2.3*), suggesting that this model only explain about nine and a half percent of the estimated wage in 2015 (*Appendix 3: Table 2.3*). The low R^2-value could be due to more noise in the data or changes in the labour market, however we cannot rule out the possibility that it might have been a coincidence or due to the composition of individuals in the dataset.

The estimates resulting from the aggregate regression estimate that the expected returns to schooling is 0.04808, or about 4.81 percent higher wage per additional year of schooling (*Appendix 3: Table 2.4*). It should be noted however that the R^2-value is relatively low, 0.169, indicating that merely 16.9 percent of the wage can be explained by years of schooling, experience and experience squared (*Appendix 3: Table 2.4*). This low level of explanation is to be expected though, as noted by Mincer in his 1974 book on the subject (1974, p.1).

6.2 Mincer with signalling effects

Table 1.8 Mincer with dummy for degree

	Equation 2			
	n	RTS	Std. Error	R-square
1997	859	0.02999***	0.006578	0.1986
2005	942	0.02415***	0.005829	0.1724
2015	683	0.01556*	0.007400	0.1331
Total	2518	0.03357***	0.002538	0.2124

Table 1.8 Mincer with dummy for degree cont.

	Equation 2							
	n	D_Vocat	Std. Error	D_Gym	Std. Error	D_Uni	Std. Error	R-square
1997	859	0.01497	0.04901	0.08679	0.05230	0.18429**	0.06092	0.1986
2005	942	0.09423	0.04859	0.16296**	0.05339	0.27091***	0.05892	0.1724
2015	683	0.07497	0.08170	0.20693**	0.08315	0.35992***	0.08866	0.1331
Total	2484	0.16281***	0.03429	0.23324***	0.03679	0.39377***	0.04115	0.2190

Incorporating the four levels of schooling defined in section 2 (*Table 1.6*), by running *Equation 2* results in a rather drastic decrease of the estimated returns to schooling (*Appendix 3: Table 3.1-3.4)* compared with the estimates from *Equation 1* (*Appendix 3: Table 2.1-2.4*). This result is quite expected as years of schooling correlates with the variables for degree, thus capturing part of the relation previously only captured in the Schooling variable. As the correlation exists regardless of whether the degree variables are included in the regression, they are essentially omitted variables in the original Mincer equation (*Equation 1*) and would be expected to capture part of the unexplained variance in the expected wage once included. The reference group in *Equation 2* is the individuals with a basic compulsory education and the coefficients for the other dummy variables represents the additional gain in expected wage in percent for each corresponding degree obtained. The results are in accordance with what is to be expected, each additional level of education correlates with an increase in the expected wage, with university degree having the largest return (*Appendix 3: Table 3.1-3.4*). The estimates for *D_Vocat*, *D_Gym* and *D_Uni* increases from 0.0149, 0.0868 and 0.1842 respectively for 1997 to 0.0749, 0.2069 (*Appendix 3: Table 3.1*) and 0.3599 respectively for 2015 (*Appendix 3: Table 3.3*). One should take notice of the fact that the P-value for the education dummies naturally varies depending on the data set. The only education level that consistently delivers significant P-value is University (*Appendix 3: Table 3.1-3.4*), thus a University degree has a statistically significant positive correlation with expected logarithmic wage in all cases, while other degrees have proved to be both statistically significant and insignificantly correlated with the expected logarithmic wage depending on which year.

6.3 Returns to schooling and the sex of the individual

Table 1.9 Mincer with dummy for sex

	Equation 3				
	n	RTS	Std. Error	Female	R-square
1997	874	0.04429***	0.004951	-0.1565***	0.1000
2005	947	0.04189***	0.004429	-0.1604***	0.1798
2015	648	0.04045***	0.005999	-0.0724*	0.2118
Total	2518	0.04869***	0.002260	-0.1191*	0.1797

Table 1.10 Mincer on subsets divided by sex

	Equation 2,	(female set)			Equation 2,	(male set)		
	n	RTS	Std. Error	R-square	n	RTS	Std. Error	R-square
1997	426	0.02434**	0.009199	0.1758	433	0.03411***	0.009361	0.2556
2005	464	0.01578*	0.007425	0.1580	478	0.02882**	0.008740	0.2182
2015	361	0.01888	0.010251	0.1415	322	0.01294	0.010887	0.1415
Total	1251	0.03414	0.005721	0.2074	1233	0.03574	0.005957	0.2497

Table 1.10 Mincer on subsets divided by sex with dummy for degree cont.

	Equation 2,	(female set)						
	n	D_Vocat	Std. Error	D_Gym	Std. Error	D_Uni	Std. Error	R-square
1997	426	0.000654	0.07102	0.03663	0.07481	0.013168	0.07971	0.1758
2005	464	0.05402	0.06221	0.07034	0.07127	0.24760***	0.07153	0.1580
2015	361	0.06324	0.12035	0.18544	0.12340	0.33203**	0.12511	0.1415
Total	1251	0.14741**	0.04907	0.17656***	0.05330	0.37238***	0.05508	0.2074

Table 1.10 Mincer on subsets divided by sex with dummy for degree cont.

	Equation 2	(male set)						
	n	D_Vocat	Std. Error	D_Gym	Std. Error	D_Uni	Std. Error	R-square
1997	433	0.02132	0.06563	0.11682	0.07172	0.26937**	0.09160	0.2526
2005	478	0.12039	0.07233	0.19658**	0.07773	0.34257***	0.09362	0.2182
2015	322	0.08665	0.11093	0.21524	0.11247	0.40696**	0.12672	0.1415
Total	1233	0.17143***	0.04723	0.25911***	0.05072	0.44352***	0.06155	0.2497

Table 1.11 Mincer on subsets divided by sex

	Equation 1,	(female set)			Equation 1,	(male set)		
	n	RTS	Std. Error	R-square	n	RTS	Std. Error	R-square
1997	432	0.03483***	0.007413	0.1675	442	0.05255***	0.006665	0.2338
2005	467	0.03378***	0.005856	0.1264	480	0.04889***	0.006601	0.1954
2015	361	0.03907***	0.008762	0.1070	323	0.04302***	0.008191	0.0953
Total	1260	0.03357***	0.002538	0.1762	1245	0.06347***	0.004325	0.2174

The returns to schooling remains approximately the same for the dataset when adding a dummy variable denoting the sex of the individual to the Mincer equation (*Equation 3*) (*Appendix 3: Table 4.1-4.4*). The *Female* coefficient for the 1997 dataset suggest that females are expected to earn 15.63 percent (*Appendix 3: Table 4.1)* less than a man. The coefficient for the dummy variable *Female* for the 2005 set is -0.1604 (*Appendix 3: Table 4.2*), which is to be interpreted as a 16 percent decrease in the expected wage for a woman. For the 2015 set the *Female* variable is now estimated to -0.0724 (*Appendix 3: Table 4.3*).

Equation 3 performed on the aggregated dataset generated similar results to *Equation 1* in regard to the returns to schooling, now 4.87 percent (*Appendix 3: Table 4.4*). The difference now being that a woman is expected to earn 11.9 percent less than a man, given the same amount of schooling and experience (*Appendix 3: Table 4.4*). This model has a R^2-value of 0.179 (*Appendix 3:* Table 2.4), a slight increase from before.

Additionally, *Equation* 1 was applied on the subsets based on sex to see whether there were any differences in the return to schooling between the sexes. According to the results there is a stronger correlation between years of schooling and expected wage for men than for women (*Appendix 3: Table 5.1- 5.8*). This is manifested by the higher R^2-values for the regressions performed on the male datasets (*Appendix 3: Table 5.1- 5.8*). Applying *Equation 1* on the female subset generates a lower return to schooling compared to that for the male subset by almost two percentage units, males having a coefficient of 5.25 percent (*Appendix 3: Table 5.4*) and females a coefficient of 3.48 percent (*Appendix 3: Table 5.3*). The development of the coefficients appears to be somewhat similar for both the sexes, remaining largely stable over time (*Appendix 3: Table 5.1- 5.8*).

Appendix 3: Table 6.1-6.8 shows the results from applying *Equation 2* on the datasets that were divided into subsets based on sex. These results are similar to those obtained when applying *Equation 1* to the same subsets (*Appendix 3: Table 5.1-5.6)* in that the estimated coefficients for females are smaller than the corresponding estimates for males (*Appendix 3: Table 6.1-6.6*). There appears to be a downward trend regarding returns to schooling for both males and females (*Appendix 3: Table 6.1-6.6)*, although the returns to schooling variable is not statistically significant for either sex in the 2015 dataset (*Appendix 3: Table 6.5-6.6*). Furthermore, the confidence intervals are large, making it difficult to draw any definitive conclusions about potential trends. The estimates for the dummy variables for degree differs between the sexes and years, for females the correlation between obtaining a University degree and wage is only statistically significant for the 2005, 2015 and aggregated dataset (*Appendix 3: Table 6.1, 6.3, 6.5 and 6.7)*. For males the University degree coefficient is both higher than the corresponding coefficient for the female subset and significant in all cases (*Appendix 3: Table 6.1-6.8*).

The resulting estimates from the Mincer equation (*Equation 1*) suggests that there is a positive correlation between years of schooling and the expected wage (*Appendix 3: Table 2.1-2.4*). The estimates vary somewhat over the different datasets which is to be expected. However, a trend cannot be statistically proven due to the small size of the difference and the vastly overlapping confidence intervals (*Appendix 3: Table 2.1-2.3*). Interestingly the size of the relative decrease in the returns to schooling estimates over time increase when degree variables are added, though it is still not possible to statistically support a trend. Furthermore, the results show a statistically significant correlation for the coefficient to the University degree dummy variable (*Appendix 3: Table 3.1-3.4*), suggesting that there might be some signalling effects present, this cannot be proven though as it likely also correlates with individual ability and other individual traits. The results also indicate different coefficients to returns to schooling when running the regression on the subsets of males and females. This suggests that the sex of the individual could potentially affect the returns to schooling for the individual. These regressions also return higher R-squared values for the regression on the subset of males, than for females, indicating that the model might better explain the variation in returns to schooling for men than women.

7. Discussion

In order to obtain real use from our results they have to be placed in a policy context, potential cautions with the models need to be assessed and the coefficients need to be interpreted. In the following section the results from the coefficients will be discussed and placed in a policy making context. Furthermore, the model, the Mincer equation and the problem with ability bias will be assessed and the effect it might have on our results will be discussed to give the reader a transparent view of the results and context that is valuable when assessing the estimated coefficients.

7.1 Signalling effect

Our results indicate that there could potentially be signalling effects affecting the expected wages. The regressions also suggest that the payoff on a University degree have increased during the period, from an 18 percent increase of the expected wage at the start of the period to a 36 percent increase at the end of the period due to obtaining a degree. A similar apparent trend can also be observed for the other degree variables, but they are not statistically significant in all cases and should thus be viewed with caution. We can observe a statistically significant coefficient for the dummy variable for University degree in all data sets. From the estimates based on the regression we also find that the coefficient for the dummy variable indicating a gymnasium degree, *D_Gym*, is also significant in all but one data set. This would suggest that there could be a correlation between obtaining a University degree above the years of schooling involved and that there could be a similar effect for obtaining a gymnasium degree. However, these results could very well be due to different underlying factors like selection effects due to innate ability, implying that those with a degree are more able and thus more productive than the individuals who does not obtain a degree, if this is the case these individuals will logically have a higher wage. It may also be due to differences in motivation among individuals which would then have a similar effect as ability differences. Hence there is no way to confirm whether there is a causal relationship between obtaining a degree, and thus signalling productivity as opposed to the relationship being rooted in ability differences among individuals.

Further this finding gives some weight to signalling theory which suggests that the degree is an important signal of productivity to the employer.

7.2 Mincer equation

The Mincer equation utilized in this paper is commonly used but have naturally also been criticised on various levels. First of all, it is a relatively simple regression analysis and, in its unaltered standard form (*Equation 1*) it does not include any measure of the innate ability of the individual resulting in the aforementioned ability bias, adding a measure of ability has empirically been shown to decrease the estimated returns to schooling with as much as 25 percent (Kjellström, 1997, p.20). Thus, part of the low R^2-values in our regressions is likely due to there being many other explanatory variables left out of the regressions. For example, the standard Mincer equation (*Equation 1*) also excludes variables with information on the ratio of students in different types of programmes, which likely affect the returns to schooling as for example the payoff to a general programme likely differs from that to a technical one and the ratio thus affect the general returns to schooling (Kjellström, 1997, p. 18). The labour market conditions also differ between different regions and professions which if it holds true will also affect the general estimated returns to schooling as it changes the wages offered. As such the standard Mincer equation does not capture all relevant variables.

Apart from the aforementioned models, a regression with some of the available variables left out in *Equation 1* was made in order to try to study how it affected the R^2 value. This new equation accounted for the sex, degree, marital status, residential area and the region in Sweden in which the person lives. Despite this the model did not receive an R^2-value much larger than 0.25, simply meaning that there is still 75 percent unexplained variance in the model. However, it should be noted that many of these new dummies are insignificant.

Kjellström and Björklund finds that the Mincer equation describes the private returns on investment in schooling rather well, as the coefficient for *Schooling* is highly related to the individuals marginal rate of return to the educational investment given that six assumptions are satisfied (2002, p.1). The aforementioned six assumptions are: “The earnings measure capture the full benefits of the investment” (Kjellström & Björklund, 2002, p.2), “The only costs of schooling are foregone earnings” (Kjellström & Björklund, 2002, p.2), “The earnings function is separable in ∂ and x so that $\frac{\partial \ln y}{\partial s}$ is independent of years of work experience”[18] (Kjellström & Björklund, 2002, p.2), “The length of working life is independent of length of schooling” (Kjellström & Björklund, 2002, p.2), “Schooling precedes work” (Kjellström & Björklund, 2002, p.2) and “The economy is in a steady state without any wage and productivity growth” (Kjellström & Björklund, 2002, p.3). They conclude that there is some effect of relaxing these assumptions that causes some results to be misleading, but for simple

[18] Where x denote work experience, y denote the expected logarithmic wage and s denote years of schooling (Kjellström & Björklund, 2002, p.2)

analyses it is a feasible method (Kjellström and Björklund, 2002, p.15). An important flaw they highlight in the Mincer equation is that it gives a linear relationship between schooling and expected logarithmic wage, thus implying that all years of schooling gives the same payoff which empirically does not seem to be the case (Björklund and Kjellström, 2002, p.15). Furthermore, as Nordin stated, the addition of degree dummies, to capture potential signalling effects changes the dynamics of the equation (Nordin, 2008, p. 2). When allowing for signalling effects in the regression we receive a model that contradicts the original linear relationship between wage and years of schooling, and instead we get a model with four different intercepts depending on the highest obtained education.

7.3 Unemployment

This thesis has a restricted scope and focus solely on the correlation and suspected positive effect between schooling and the expected logarithmic wage. Thus, it has not studied any potential effects schooling might have on unemployment, nor is this captured by any model as we have restricted our datasets to only individuals with employment and thus a wage. Even though these potential effects are not studied in this paper they likely correlate, especially in a labour market increasingly affected by changes in production technology as well as changing spatial concentrations of production and globalisation. It could thus be a satisfactory subject for further research.

This unemployment factor is likely to affect the result in both ways. By excluding the unemployed observations in our data, and since there is a strong correlation between schooling and employment, the excluded observations likely have less years of schooling and likely also have lower level of finished education (Galte Schermer, 2018). Thus, suggesting that the there is a larger payoff to more education than suggested by the regression. The potential overestimating effect would appear since the tests are based on the assumption that education is expected to lead to employment while there are unemployed individuals that are educated. It is difficult to determine exactly how this will affect our results and to what extent this is a problem.

It should be noted that using the zero workhours as a restriction will allow people working jobs of lower qualification and thus likely includes individuals with a productivity above what is represented by their hourly wage. However, this could happen with a restriction based upon an age span as well and is furthermore not likely to be a significant cause of error.

7.4 Ability bias

Among the most prominent problems with estimates of returns to schooling is ability bias, i.e. that the wage reflects the individual's productivity and thus incorporate both effects from schooling and experience but also the innate ability of the individual. While the regression captures the effects of experience and schooling it does not incorporate any measure of ability as it is notoriously difficult to measure properly, thus resulting in an omitted variable bias. At the same time individuals with higher ability are more likely to acquire more schooling (Blackburn & Neumark, 1993, p. 5), according to human capital theory, as they would then have lower marginal costs of acquiring schooling compared to their peers (Rospigliosi et al, 2014, p.3). This aspect of ability is difficult to capture in the data and will likely also cause an uneven distribution of the ability among the people who acquire further schooling. Due to limitations in our data set we do not have a proper way to control for these effects, as we do not have data on the individual's innate ability.

The effect of the selection bias part of the overarching ability bias could arguably be smaller in the context of Sweden though since it to a large degree stems from the assumption that more able individuals will have lower marginal costs of acquiring schooling and thus acquire more of it. As schooling is free in Sweden these marginal costs are lower for everyone, thus reducing this threshold. Hence it could be argued that the specific portion of the ability bias would be on the lower end of the spectrum. As for the remaining portion of the ability bias, that more able individuals in theory should earn more than less able individuals given the same amount of schooling it likely manifests itself in the results as an upwards bias in the coefficients. It is therefore likely that it does affect the results to an extent. We cannot assume that this effect has been constant over time, and it is thus not entirely impossible that it might affect any potential trend in returns to schooling.

7.5 Change in returns to schooling?

Although our results appear to follow the trend of decreasing returns to schooling found by Palme and Wright (1998, p.5), we cannot statistically support that our results follow such a trend due to large confidence intervals. Our 1997 value for the returns to schooling is higher than the 1991 values, both for men and women, in Palme and Wright's paper which might be due to a number of factors, including an actual rise in returns to schooling, which would be supported by the findings of Björklund et al. that concluded that the returns to schooling had risen during the time (2004, p. 45). It is also possible that the difference stems from differences in the data utilized, i.e. that the difference stems from noise in the data. Since all these studies utilize the basic Mincer equation (*Equation 1*) it appears unlikely that the econometrics utilized are the cause of error. Palme and Wright do naturally utilize another data set as a point of reference than we do, the Swedish level of Living Survey, thus the way the variables are composed, derived from the questions asked, may vary which could potentially lead to differences in the results. The difference is moderate as well, the estimated returns to schooling in 1991 being 4.0 percent for males and 3.5 percent for women while our 1997 estimate is 4.429 percent, including both groups, albeit with a negative female dummy variable. Worth noting is that our estimates fall in between these two, they are lower than what is estimated by Björklund et al. while suggesting that the trend proposed in Palme and Wright's research did not continue.

The results from the regressions confirms that there is a positive correlation between years of schooling and wage, as all the *Schooling* coefficients are statistically significant seeing that the P-value is below 0.05. However, it is difficult to draw any conclusions on the development of the rate of returns to schooling during the timespan since all the schooling coefficients are close to each other.

7.6 Policy relevance

The returns to schooling are an important instrument for constructing public policy related to schooling as it is an estimate of the individuals private economic return for each additional year of schooling. If one for example wants to make more individuals go into higher education it should, based on the human capital model, be feasible to raise the marginal returns to education, everything else constant as it will then increase the economic incentives to acquire more education. Decreasing marginal costs of schooling can also positively affect the schooling decision. While the findings in Björklund et al. differ from our own findings in that we cannot prove an increase, they also present another interesting aspect regarding the marginal costs of schooling. They find that changes in taxing and student benefits, effectively has kept the total private return to additional schooling constant (Björklund et al. 2004, p. 46). If we consider this while viewing our own results it is not unlikely that the total private returns has decreased somewhat even though the returns to schooling appears to be unchanged over the period. Further studies of this is necessary to draw a definitive conclusion though.

The findings in this study suggest that the amount of education obtained is not the only variable to regard in this matter, but that the highest degree obtained is highly relevant as well. This is supported by the large decline in the estimated returns to schooling in the regression with degree dummy variables, suggesting that having a degree might explain a relatively large portion of the expected wage. Thus, implying an importance of obtaining a degree in line with the signalling theory model. As such these results could have policy implications as it implies that it might be as important to obtain a degree as it is to obtain additional years of schooling. Should that hold true it could suggest policies aiming at raising the ratio of students who finish their studies would be highly beneficial.

As stated by Björklund et al. there are more aspects of the total private returns to education than the returns to schooling, i.e. the wage premia obtained for one additional year of schooling, there is also the cost of schooling aspect which has been addressed briefly above (Björklund et al. 2004, p.46). From a policy standpoint the aim is usually to raise the incentives to acquire further schooling. Thus, policymakers have two sets of tools they can use, one set to adjust the costs of acquiring more schooling and one set to try and raise the returns to schooling. As the returns to schooling are largely dependent upon the wage structure in the market this is arguably harder to achieve than lowering the costs which are, in the Swedish case largely dependent on state provided subsidies and loans. Hence, increasing the returns to schooling values might be difficult, however viewing our results it is suggested that the individual's returns increase a lot if they obtain a degree and thus policy solutions that help individuals to finish their studies could potentially raise the incentives to obtain further schooling. The incentives may increase in accordance with both models as they both assume individuals acquire schooling until the marginal costs of further schooling is equal to the marginal benefits of doing so (Rospigliosi et al. 2014, p.3). The significant degree coefficients in our regressions (*Appendix 3: Table 3.1-3.4*) would suggest that there could be signalling values present and that they would then have a significant correlation with the wage premia for the individual. This indicates that policy evaluation on the subject that only apply a human capital theory approach could be misleading due to the exclusion of potential signalling effects. In accordance with signalling theory the wage premia are hinged upon the assumption that obtaining a degree is associated with a certain cost for the individual such that the cost is lower for productive individuals, since this is what makes the degree a credible signalling instrument (Rospigliosi et al. 2014, p.4). Hence, policies aimed at helping individuals finish their studies could potentially lower the wage premia from obtaining a degree as it is no longer such a strong indicator of productivity.

7.7 Differences between sexes

The difference regarding returns to schooling between the sexes does not necessarily have to be a consequence of discrimination towards women. One of the weaknesses with the Mincer equation is that it fails to capture that the sort of education acquired will have a substantial impact on the future wage. There are some notable differences between the sexes regarding the choice of education in Sweden, men tend to work in high-wage sectors to a larger extent compared to women (Statistics Sweden, 2016). This theory could to an extent be supported by the difference in coefficients to the variable *D_Uni,* between the male and female subset (*Appendix 3: Table 6.1-6.*4) suggesting that the degree results in different returns depending on the sex of the individual. This is however hinged upon the assumption that the difference might be due to the degrees obtained on an aggregate level are aimed towards different sectors of the labour market in the male and female subset. Should the degrees be equal on an aggregate level this no longer holds true. Even if this is a possible source for some of the differences between men and women, previous research from Statistics Sweden suggests that, when the sector differences have been accounted for, a woman earns 94 percent of what a man earns (Statistics Sweden, 2016). Thus, there is still reason to believe that there exists some wage discrimination towards women. This could potentially be explained via the externalities in signalling theory. As the sex of the individual is usually stated as unalterable it is one of the indices of the individual (Spence, 1973, p.3). The employer views both signals and indices when making a decision and in this context the wage is thus determined both by the signal of educational attainment and the indices, among them the sex of the participant (Spence, 1973, p.16). As the externalities are present the individual will be assessed in part based on the group average, in this case for males and females, and any differences in educational attainment or productivity between the groups will affect the individual to an extent (Spence, 1973, p.16). Thus, it is possible that differences in the market data regarding the groups could be affecting the wage and also the returns for obtaining a degree.

In regard to differences based upon the sex of the individual the returns to schooling results are similar to those obtained by Palme and Wright that concluded that the males appear to have a higher rate of return (Palme & Wright, 1998, p.6). The results from applying *Equation 2* to the subsets of males and females for each respective dataset (*Appendix 3: Table 6.1-*

6.6) are similar to those obtained when applying *Equation 1* to the same subsets (*Appendix 3: Table 5.1-5.6)* in that the estimated coefficients for females are smaller than the corresponding estimates for males. However, one should note that we cannot draw any conclusions from the 2015 estimates as these are statistically insignificant for both males and females (*Appendix 3: Table 6.5-6.6)*. Furthermore, the estimated coefficients for *D_Uni*, representing obtaining a University degree are smaller for females than males in both 1997 and 2005 (*Appendix 3: Table 6.1-6.4*).

8. Conclusion

This paper has estimated the returns to schooling in Sweden based on three datasets from 1997, 2005 and 2015 respectively. The resulting estimates were rather equal with large confidence intervals (*Appendix 3: Table 2.2-2.4*); thus, no conclusive proof of a trend could be established. To an extent this contradict the previous research conducted by Nordin that suggested a trend of increasing returns to schooling in Sweden during the late 90s and early 2000s (2008, p.8) and Björklund et al. that indicated an upward trend in returns to schooling in Sweden during the late half of the 90s (2004, p.45).

The estimates from the regressions with dummy variables for the highest degree of education obtained resulted in significant positive coefficients for the University degree dummy variable, *D_Uni* in all cases (*Appendix 3: Table 3.1-3.3)* including the regressions on the subsets for males and females *(Appendix 3: Table 6.1-6.8*). This suggests that there could be signalling effects present, however, these could be affected by other factors correlating with the attainment of a degree, notably ability. Thus, no definitive conclusion can be drawn about the presence of signalling effects although their potential presence is implied.

The regressions on the separate subsets for males and females resulted in statistically significant estimates of the coefficients for *Schooling* that consistently differed for the two subsets. Thus, implying that there is some difference in the returns to schooling for males and females. However, it should be noted that the confidence intervals for these estimates are large and thus the difference cannot be supported statistically. These supposed differences need not be due to discrimination and could be due to different field of study being chosen. However, such differences in aggregate educational attainment within the groups could, in accordance with signalling theory affect the wage of an individual within the group as the individual is assessed based on the average member of their group.

9. References

Björklund, A. Edin, P. Fredriksson, P. Krueger, A. (2004) Education, equality and efficiency - An analysis of Swedish school reforms during the 1990s [PDF] Available online: https://www.ifau.se/globalassets/pdf/se/2004/r04-01.pdf [Accessed 04 May 2018]

Björklund, A & Kjellström, C. (2002) Estimating the return to investments in education: how useful is the standard Mincer equation?, *Economics of Education Review, vol. 21, no. 3, pp.195-210*, Available online: http://resolver.ebscohost.com/openurl?sid=EBSCO%3aecn&genre=article&issn=02727757&isbn=&volume=21&issue=3&date=20020601&spage=195&pages=195-210&title=Economics+of+Education+Review&atitle=Estimating+the+Return+to+Investments+in+Education%3a+How+Useful+Is+the+Standard+Mincer+Equation%3f&btitle=Economics+of+Education+Review&jtitle=Economics+of+Education+Review&series=&aulast=Bjorklund%2c+Anders&id=DOI%3a&site=ftf-live [Accessed 11 May 2018]

Blackburn, M.L. & Neumark, D. (1993). Omitted Ability Bias and the Increase in Returns to Schooling, *Journal of Labour Economics, vol. 11, no. 3, pp.521-544*, Available Online: http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&sid=ac067918-07ad-4685-97be-439e06ebc67f%40sessionmgr104 [Accessed 02 May 2018]

Card, D. (2001). Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica, vol. 69, no. 5, pp.1127-11260*, Available online: https://www.econometricsociety.org/publications/econometrica/2001/09/01/estimating-return-schooling-progress-some-persistent [Accessed 02 May 2018]

Carlgren, F. (2017), Real löneutveckling i Sverige. Available online: https://www.ekonomifakta.se/Fakta/Arbetsmarknad/Loner/Loneutveckling-och-inflation/ [Accessed 14 May 2018]

Galte Schermer, I. (2018) Arbetslöshet efter utbildningsnivå. Available online: https://www.ekonomifakta.se/Fakta/Arbetsmarknad/Arbetsloshet/Arbetsloshet-efter-utbildningsniva/ [Accessed 07 May 2018]

GESIS Data Archive for the Social Sciences (2005) ZA4350_backgroundvar_se [PDF] Available at: https://dbk.gesis.org/dbksearch/download.asp?db=E&id=50776 [Accessed 20 April 2018]

GESIS Data Archive for the Social Sciences (2015) ZA6770_backgroundvar_se [PDF] Available at: https://dbk.gesis.org/DBKsearch/download.asp?db=E&id=61769 [Accessed 02 May 2018]

GESIS Data Archive for the Social Sciences (2015) ISSP 2015 Work Orientations IV - ZA6770 - Variable Report [PDF] Available at: https://dbk.gesis.org/DBKsearch/download.asp?db=E&id=62773 [Accessed 02 May 2018]

ISSP Research Group (1999): International Social Survey Programme: Work Orientations II - ISSP 1997. GESIS Data Archive, Cologne. ZA3090 Data file Version 1.0.0, doi:10.4232/1.3090 Accessed via: https://dbk.gesis.org/dbksearch/sdesc2.asp?no=3090&search=issp%201997&search2=&field=all&field2=&DB=e&tab=0&notabs=&nf=1&af=&ll=10 [Accessed 24 May 2018]

ISSP Research Group (2013): International Social Survey Programme: Work Orientation III - ISSP 2005. GESIS Data Archive, Cologne. ZA4350 Data file Version 2.0.0, doi:10.4232/1.11648, Accessed via: https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4350&search=issp%202005&search2=&field=all&field2=&DB=e&tab=0&notabs=&nf=1&af=&ll=10 [Accessed 24 May 2018]

ISSP Research Group (2017): International Social Survey Programme: Work Orientations IV - ISSP 2015. GESIS Data Archive, Cologne. ZA6770 Data file Version 2.1.0, doi:10.4232/1.12848, Accessed via: https://dbk.gesis.org/DBKsearch/SDESC2.asp?no=6770&tab=3&db=E [Accessed 24 May 2018]

Kjellström, C. (1997) Omitted ability bias and the wage premium for schooling: new Swedish evidence, Stockholm: Univ., Institutet för social forskning

Mincer, J. (1974), Schooling, experience, and earnings, New York: National Bureau of Economic Research

Nordin, M (2008) Ability and Rates of Return Schooling-Making Use of the Swedish Enlistment Battery Test, *Journal of Population Economics, vol. 21, no. 3, pp.703-717,* Available online: http://eds.b.ebscohost.com/eds/Citations/FullTextLinkClick?sid=474edd79-ff77-4695-8538-44604447a71d@sessionmgr101&vid=0&id=pdfFullText [Accessed 23 April 2018]

OECD (2017) Education at a Glance 2017 [PDF] Available at: https://www.oecd-ilibrary.org/docserver/eag-2017-en.pdf?expires=1526292554&id=id&accname=ocid177253&checksum=C448394D55BF81F976B8421EA8DB6F96 [Accessed 14 May 2018]

Palme, M. & Wright, R. (1998) Changes in the rate of return to education in Sweden: 1968-1991, *Applied Economics, vol. 30, no. 12, pp.1653-1663,* Available online: http://eds.b.ebscohost.com/eds/viewarticle/render?data=dGJyMPPp44rp2%2fdV0%2bnjisfk5Ie46bFRr6iySrek63nn5Kx94um%2bTK2osUewpq9Qnq%2b4SrSwr0met8s%2b8ujfhvHX4Yzn5eyB4rOvSbatsUmzqbNMpOLfhuWz44ak2uBV4OrmPvLX5VW%2fxKR57LOvTLKttlCxnOSH8OPfjLvm4n7E6%2bqE0tv2jAAA&vid=1&sid=a32f956e-f077-4417-b4c9-3cdd25efed1c@sessionmgr104 [Accessed 23 April 2018]

Rospigliosi, A.P., Greener, S. Bourner, T. and Sheehan, M, (2014) Human capital or signalling unpacking the graduate premium, *International Journal of Social Economics, vol. 41, no. 5, pp.420-432*, Available online: http://web.a.ebscohost.com.ludwig.lub.lu.se/ehost/viewarticle/render?data=dGJyMPPp44rp2%2fdV0%2bnjisfk5Ie46bFRr6iySrek63nn5Kx95uXxjL6rrU%2btqK5Jtpa0UrKpuEi0lr9lpOrweezp33vy3%2b2G59q7S7Oqrk%2b3rbU%2b6tfsf7vb7D7i2Lt94eSkjN%2fdu1nMnN%2bGu6eyTLWrrkuk3O2K69fyVeTr6oTy2%2faM&vid=2&sid=79c14964-2c61-4bae-adf5-ac538651a660@sessionmgr4008 [Accessed 07 May 2018]

Statistics Sweden (2003). A-regioner 2006-01-01 [PDF] Available online: https://www.scb.se/Grupp/Produkter_Tjanster/Verktyg/_Dokument/A_regiontxt.pdf [Accessed 02 May 2018]

Statistics Sweden (2016) Kvinnor har lägre lön, inkomst och pension än män. Available online: https://www.scb.se/hitta-statistik/statistik-efter-amne/levnadsforhallanden/jamstalldhet/jamstalldhetsstatistik/pong/statistiknyhet/pa-tal-om-kvinnor-och-man-2016/ [Accessed 23 May 2018]

Statistics Sweden (2017) Befolkningens utbildning 2016 [PDF] Available online: https://www.scb.se/contentassets/66fa9c1d1f904b4aa22ac4a816d9e9a5/uf0506_2016a01_sm_uf37sm1701.pdf [Accessed 24 April 2018]

Skolverket, (n.d.) Gymnasieskola, Available online: https://www.skolverket.se/skolformer/gymnasieutbildning/gymnasieskola [Accessed 14 May 2018]

Skolverket (2015) Grundskola och Fritidshem. Available online: https://www.skolverket.se/skolutveckling/larande/nyanlandas-larande/information-till-nyanlanda-elever/grundskola-och-fritidshem-1.236748 [Accessed 22 May 2018]

Spence, A.M. (1973) Job Market Signaling, *Quarterly Journal of Economics, vol. 87, no. 3, pp. 355-374*, Available online: http://web.a.ebscohost.com/ehost/viewarticle/render?data=dGJyMPPp44rp2%2fdV0%2bnjisfk5Ie46bFRr6iySrek63nn5Kx95uXxjL6qrUqypbBIr66eSa%2bwrkq4qbU4v8OkjPDX7Ivf2fKB7eTnfLujr0%2b1r69Mr6izS6Ti34bls%2bOGpNrgVePZ7D7y1%2bVVv8Skeeyzrki1qrdLtpzkh%2fDj34y73POE6urjkPIA&vid=7&sid=dfd3aecf-91f7-403a-8a71-aab81a10d112@sessionmgr4008 [Accessed 24 April 2018]

Umeå Universitet (n.d.) Available online: http://www.umu.se/forskning/stark-forskning/starka-forskningsmiljoer/forskningsdatabasvisning/?code=345&currentView=description [Accessed 23 April 2018]

Uppsala Universitet, (n.d.) Available online: https://www.uu.se/student/examen/utbildningsnivaer_examina/ [Accessed 14 May 2018]

Zentralarchiv für Empirische Sozialforschung (2001) ISSP 1997 Work Orientations II - ZA3090 - Codebook [PDF] Available at: https://dbk.gesis.org/dbksearch/download.asp?db=E&id=15666 [Accessed 20 April 2018]

Appendix 1: Variable List

Background Variables:

educyrs (Work Orientations II, 1997)
The variable used to describe years of schooling in our mincer equation from the 1997 dataset is named educyrs, a shorthand for education years measures the individuals years in school. It is worth noting that even non completed full time education is included here (Zentralarchiv für Empirische Sozialforschung, 2001, s. 48).

wrkhrs (Work Orientations II, 1997)
The name is short for work hours and the variable describes the number of hours an individual works weekly (Zentralarchiv für Empirische Sozialforschung, 2001, s. 266). We use this variable in order to obtain a comparable measurement of an individual's income as just monthly income gives a skewed representation of return to schooling.

s_rinc (Work Orientations II, 1997)
The variable measures the respondents monthly income measured in thousands of Swedish Kroners (SEK) (Zentralarchiv für Empirische Sozialforschung, 2001, p. 195). To make it easier to work with the data we multiplied each entry by 1000 so that it is instead measured in Swedish Kroner instead of in Thousands of Swedish Kroner.

s_degr (Work Orientations II, 1997)
This variable contains information on the educational degree obtained by the respondent, here only corresponding to degrees in Sweden (Zentralarchiv für Empirische Sozialforschung, 2001, p. 266). It is divided into the following categories: "Primary or comprehensive school", "Vocational school (post 1972)", "Vocational school (pre 1972)", "Folk high-school", "Lower secondary school", "3 or 4 year gymnasium", "Higher secondary school", "University studies without deg" and "University degree" (Zentralarchiv für Empirische Sozialforschung, 2001, p. 171).

Sex (Work Orientations II, 1997)
The variable contains the gender of the individual in the observation based on register data from the country in question (Zentralarchiv für Empirische Sozialforschung, 2001, p.80)

WRKHRS (Work Orientations III, 2005)
The work hours variable is made of the survey question "How many hours do you usually work per week?" upon which the respondent enter an integer of hours (ISSP, 2005, p.9). The lowest recorded measurement being 2, the highest 80 and the individuals responding 0 have been classified as currently not in the labour force (ISSP, 2005, p.9).

SE_RINC (Work Orientations III, 2005)
The income variable, in the dataset named SE_RINC measures the individuals monthly earnings based on the question: "What is your approximate monthly income before taxes? ………….. SEK per month." (ISSP, 2005, p. 19). The ISSP further states that the variable has been allowed to vary between 0 and 998000, this being as 999999 codes for No answer (ISSP, 2005, p.19).

EDUCYRS (Work Orientations III, 2005)

The years of schooling variable in the 2005 dataset is based upon the results from the question "In total, how many years have you been in fulltime education? Count from the low stadium of the comprehensive school ……. years" (ISSP, 2005). The variable has been defined so that it is allowed to take on values between zero and twenty five, and the ISSP further note that individuals usually enter school at six or seven years old (2005).

DEGREE (Work Orientations III, 2005)
The degree variable from the 2005 dataset corresponds to the question "What is your highest level of education at present? Please indicate one alternative only", with the alternatives: "Primary or comprehensiveschool", "Vocational school 2 years (1972-1992)", "Vocational school (post 1992)", "Vocational school (pre 1972)", "Alternative secondary school", "Lower secondary school", "3 or 4 year gymnasium (academic track pre 1995)", "Higher secondary school", "University studies without degree" and "University degree" (ISSP, 2005, p.6). It is worth noting here that some differences arise in the alternatives if compared to the 1997 survey. Notably the 1997 alternative "Folk high-school" correspond to the alternative "Alternative secondary school", both representing the Swedish Folkhögskola. Further the alternative "Vocational school post 1972" from the 1997 dataset has been replaced by the two alternatives "Vocational School 2 years (1972-1992)" and "Vocational School (post 1992)".

SEX (Work Orientations III)
A variable on the sex of the responent, the question is not asked to the respondent but rather fetched from a national register (ISSP, 2005, p.1).

URBRURAL (Work Orintations III)
A regional variable based on the question: "Is the place where You live…" (ISSP, 2005, p. 33) followed by alternatives "A big city", "A suburb or in the outskirt of a big city", "A town or small city", "A country village" and "A farm or home in the country" (ISSP. 2005, p. 33).

WRKHRS (Work Orientations IV)
The work hours variable has not changed significantly since the 2005 dataset, there have however been a few minor changes in the question which now goes: "How many hours, on average, do you work for pay in a normal week? (Please include, if any, overtime)" (GESIS Data Archive for the Social Sciences, 2015, p.11). The main factor that changed is the addition of "for pay" in the question. This change could have some effect on people's answers although on aggregate it should remain roughly equal.

SE_RINC (Work Orientations IV)
SE_RINC is, as in previous datasets personal income, it corresponds to the question: "On average, what is your monthly income before taxes? On average my income before taxes is ……. SEK/month"(GESIS Data Archive for the Social Sciences, 2015, p. 42). Thus this variable is rather identical to the one found in previous datasets.

EDUCYRS (Work Orientations IV)
In the 2015 dataset the education years is very similar to the previous instances. However the question asked has been modified slightly and now goes: "How many years of full-time education have you had? include primary school, and, if applicable, gymnasium and college/university education. *If you are currently in education, count the number of years you have completed so far."*(GESIS Data Archive for the Social Sciences, 2015, p.5) In essence the changes are just clarifications of the previously asked questions and should not to a noteworthy degree affect our results.

DEGREE (Work Orientations IV)
The DEGREE variable in the 2015 measures highest degree obtained in education based on the following alternatives: “Primary school”, “Lower secondary (secondary completed does not allow entry to university: obligatory school)”, “Upper secondary (programs that allows entry to university”, “Post secondary, non-tertiary (other upper secondary programs toward labour market or technical formation)”, “Lower level tertiary, first stage (also technical schools at tertiary level)”, “Upper level tertiary (Master, Doctor)” and “No answer”.

URBRURAL
The variable holds information on the region in which the respondent is residing, based on the following question: “The place where you live, is it…” (GESIS Data Archive for the Social Sciences, 2015, p. 47) followed by these alternatives: “A big city”, “The suburbs or outskirts of a big city”, “A town or small city”, “A country village” and “A farm or home on the country” (GESIS Data Archive for the Social Sciences, 2015, p. 47)

SE_REG (Work Orientations IV
This regional variable is based upon the following 70 Swedish a-regions (GESIS Data Archive for the Social Sciences, 2015 (CODEBOOK), p. 843): “Stockholm/Södertälje”, “Norrtälje”, “Enköping”, “Uppsala”, “Nyköping”, “Katrineholm”, “Eskilstuna”, “Mjölby/Motala”, “Linköping”, “Norrköping”, “Jönköping”, “Tranås”, “Enköping/Nässjö/Vetlanda”, “Värnamo”, “Ljungby”, “Växjö”, “Västervik”, “Hultsfred/Vimmerby”, “Oskarshamn”, “Kalmar/Nybro”, “Visby”, “Karlskrona”, “Karlshamn”, “Kristianstad”, “Hässleholm”, “Ängelholm”, “Helsingborg/Landskrona”, “Malmö/Lund/Trelleborg”, “Ystad/Simrishamn”, “Eslöv”, “Halmstad”, “Falkenberg/Varberg”, “Göteborg/Allingsås”, “Uddevalla”, “Trollhättan/Vänersborg”, “Borås”, “Lidköping/Skara”, “Falköping”, “Skövde”, “Mariestad”, “Kristinehamn/Filipstad”, “Karlstad”, “Säffle/Åmål”, “Arvika”, “Örebro”, “Karlskoga”, “Lindesberg”, “Västerås”, “Köping”, “Fagersta”, “Sala”, “Borlänge/Falun”, “Avesta/Hedemora”, “Ludvika”, “Mora”, “Gävle/Söderhamn”, “Hudiksvall/Ljusdal”, “Sundsvall”, “Härnösand/Kramfors”, “Sollefteå”, “Örnsköldsvik”, “Östersund”, “Umeå”, “Skellefteå”, “Lycksele”, “Piteå”, “Luleå/Boden”, “Haparanda/Kalix” and “Kiruna/Gällivare”. (GESIS Data Archive for the Social Sciences, 2015, p.48-50). The a-region classification was made in the 1960s as a form of, then, industry geographically coherent regions (SCB, 2003, p. 89).

SEX (Work Orientations IV)
This variable states the sex of the individual, the individual is given the choices: “Male”, “Female”, and not to answer (GESIS Data Archive for the Social Sciences, 2015, p.1)

SEX (Aggregated dataset)
The variable sex consists of the entries from the variable SEX in the three ISSP Work Orientations Datasets.

Age (Aggregated dataset)
The variable *Age* in the aggregated dataset consists of the entries from the age variables from each ISSP dataset (Work Orientations II, Work Orientations III and Work Orientations IV), and denotes the age of the individual.

Schooling (Aggregated dataset)
In the merged dataset Schooling represents the individuals years of schooling. The variable aggregates the entries from the educyrs variables from each corresponding ISSP Work Orientations dataset.

Degree (Aggregated dataset)
The Degree variable in the merged dataset aggregates all entries from the country specific degree variables for Sweden, s_degr and SE_DEGR in ISSP Work Orientations II and Work Orientations III and IV respectively.

SE_RINC (Aggregated dataset)
SE_RINC denotes the average monthly income of the individual in the aggregated dataset. It consists of all the entries from the monthly income variables, *s_rinc (Work Orientations II), SE_RINC (Work Orientations III)* and *SE_RINC (Work Orientations IV),* in the 1997, 2005 and 2015 ISSP dataset respectively.

WRKHRS (Aggregated dataset)
WRKHRS denotes the average hours of payed work the individual conducts weekly and consists of all entries in the work hours variables, *wrkhrs (Work Orientations II), WRKHRS (Work Orientations III)* and *WRKHRS (Work Orientations IV).*

Our variables

Sex

The variable *Sex* is based upon the corresponding variable in each dataset (*sex (Work Orientations II), SEX (Work Orientations III), SEX (Work Orientations IV)* and *SEX (aggregated dataset)*).

Wage

The wage variable is based on the monthly income variable in each corresponding dataset. Thus depending on the dataset it is renamed from: *se_rinc (Work Orientations II), SE_RINC (Work Orientations III), SE_RINC (Work Orientations IV)* and *SE_RINC (Aggregated dataset)*

Hours

The variable *Hours* represents the amount of hours worked weekly given by the respondents and is thus based upon the background variables for the corresponding datasets: *wrkhrs (Work Orientations II), WRKHRS (Work Orientations III), WRKHRS (Work Orientations IV)* and *WRKHRS (Aggregated dataset).*

Whour

This variable gives the expected hourly wage in the observation. It is created by multiplying *Wage* by 12 to get the yearly wage, and divides this by the variable *Hours* multiplied by 52, that is divided by the amount of hours worked weekly times the number of weeks in a year. Thus, it is based on the following formula:

$$Whour = \frac{Wage * 12}{Hours * 52}$$

Lnwage

Represents the logarithmic expected hourly wage by taking the natural logarithm of *Whour*. Thus, it can be defined as: $Lnwage = \ln(Whour)$

Schooling

Schooling gives the number of years in schooling the respondent has reported. It is based on the years of schooling variable for the corresponding dataset: *educyrs (Work Orientations II), EDUCYRS (Work Orientations III), EDUCYRS (Work Orientations IV)* and Schooling (Aggregated dataset)

Exp

Exp denotes the experience of the individual to approximate the individuals work experience during which it can be assumed that on the job training takes place. It is defined as: $\exp = age - schooling - 6$. It is constructed in this way to capture experience on the job and thus it excludes schooling and the six years before starting school.

Exp2

Exp2 denotes the squared experience term and is included to capture the nonlinearity of the relation between expected wage and experience. It is defined as: $exp2 = exp^2$.

D_Vocat

D_Vocat is a degree based dummy variable that takes on the value 1 if the highest degree obtained by the degree is Vocational education and 0 if it is not. The highest obtained degree is defined as Vocational education if the individual has answered: “Vocational school (post 1972)”, “Vocational school (pre 1972)”, “Vocational school (1972-1992)”, “Vocational school (post 1992)”, “Vocational school” or “Vocational school 1963-70” in the *DEGREE* variable in the aggregated dataset.

D_Gym

D_Gym is a degree based dummy variable that takes on the value 1 if the highest degree obtained by the degree is Gymnasium, the Swedish equivalent of High school and 0 if it is not. The highest obtained degree is defined as Gymnasium if the individual has answered: “3 or 4-year gymnasium (academic track)”, “gymnasium (academic track post 1992)” or “higher secondary school” in the *DEGREE* variable in the aggregated dataset.

D_Uni

D_Uni is a degree based dummy variable that takes on the value 1 if the highest degree obtained by the degree is a University degree and 0 if it is not. The highest obtained degree is defined as a University degree if the individual has answered: “University degree, 3 years or more”, “doctor’s degree”, “university degree, less than 3 years” or “university degree” in the *DEGREE* variable in the aggregated dataset.

D_mar

D_mar is a dummy variable regarding the civil status of the individual, it takes on the value 1 if the individual is married and 0 if the individual is not. The individual is defined as married if he or she has answered: “Married” or “civil partnership” in the *CIVSTAT* variable in the aggregated dataset.

D_sep

D_sep is a dummy variable that is 1 if a person is separated or divorced and 0 if the person is not. An individual is defined as separated if the person has the answer “divorced” or separated in the variable CIVSTAT in the aggregated data set.

D_wid

D_wid is a dummy variable that takes the value 1 if a person is widowed and 0 if the individual is not. The individual is defined as widowed if they have answered “widowed” in the *CIVSTAT* variable in the aggregated dataset.

D_urb

D_urb is a dummy variable denoting that the respondent lives in an urban area, it is constructed based on the answer “Urban” in the variable *urbrural (Work Orientations II)*.

D_urb

D_sub is a dummy variable denoting that the respondent lives in a suburban area, it is constructed based on the answer “Suburban, city-town” in the variable *urbrural (Work Orientations II).*

D_smagot

D_smagot is a regional dummy variable denoting that the respondent is from the Småland-Gotland region, corresponding with the answer “Smaland Gotland” in the variable *s_reg (Work Orientations II).*

D_south

D_south is a regional dummy variable that takes on the value 1 if the individual is from the region south. The individual is defined as belonging to the region south if the individual has answered : “ystad /simrishamn”, “Kristianstad”, “Malmoe”, “Helsingborg/Landskrona”, “Halmstad”, “Hässleholm”, “Eslöv”, “Karlskrona” or “Karlshamn” in *SE_REG* variable in the aggregated dataset.

D_west

D_west is a regional dummy variable that takes on the value of 1 if the individual is belonging to the region west, and 0 if the individual is not. The individual is defined as belonging to the region west if he or she has answered: “Kristinehamn/Filipstad”, “Borås”, “Falköping”, “Falkenberg/Varberg”, “Lidköping/Skara”, “Seffle”, “Trollhättan/Vänersborg”, “Ludvika”, “Uddevalla”, “Goeteborg”, “Goeteborg/Allingsås” and “Karlstad” in *SE_REG* variable in the aggregated dataset.

D_northmid

D_northmid is a regional dummy variable that takes on the value of 1 if the individual is belonging to the region northmid and 0 if he or she is not. The individual is defined as belonging to the region northmid if he or she has answered: “Mora”, “Avesta/Hedemora”, “Bollnäs”, “Borlänge/Falun” or “Lycksele” in the variable *SE_REG* in the aggregated dataset

D_north

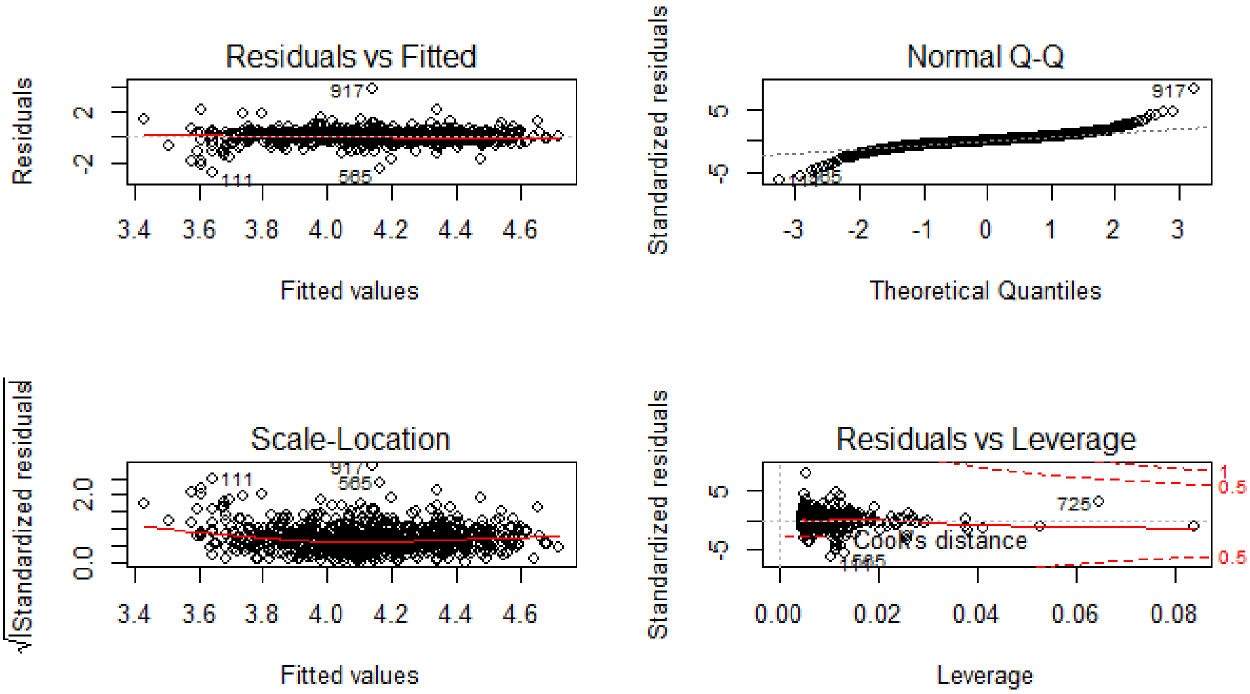
D_north is a regional dummy variable that takes on the value 1 if the individual is belonging to the region north and 0 if the individual is not. The individual is defined as belonging to the region north if the he or she has answered: “Haparanda/Kalix”, “Kiruna/Gällivare”, “Luleå/Boden”, “Piteå” and “Umeå” in the *SE_REG* in the aggregated dataset.

D_stock

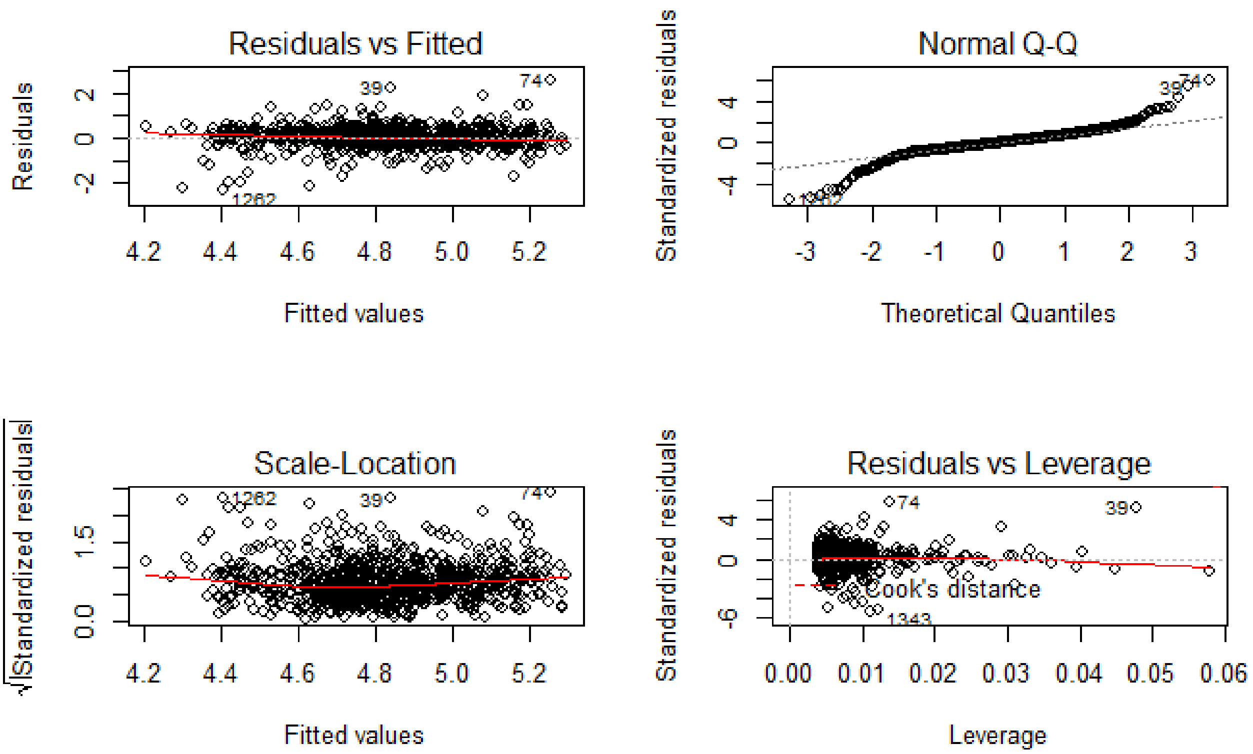
D_stock is a regional dummy variable that takes on the value 1 if the region is belonging to the region Stockholm and 0 if the individual is not. An individual is defined as belonging to the region Stockholm if he or she has answered “Stockholm/Södertälje” in the SE_REG in the aggregated dataset.

Appendix 2: Econometric validation

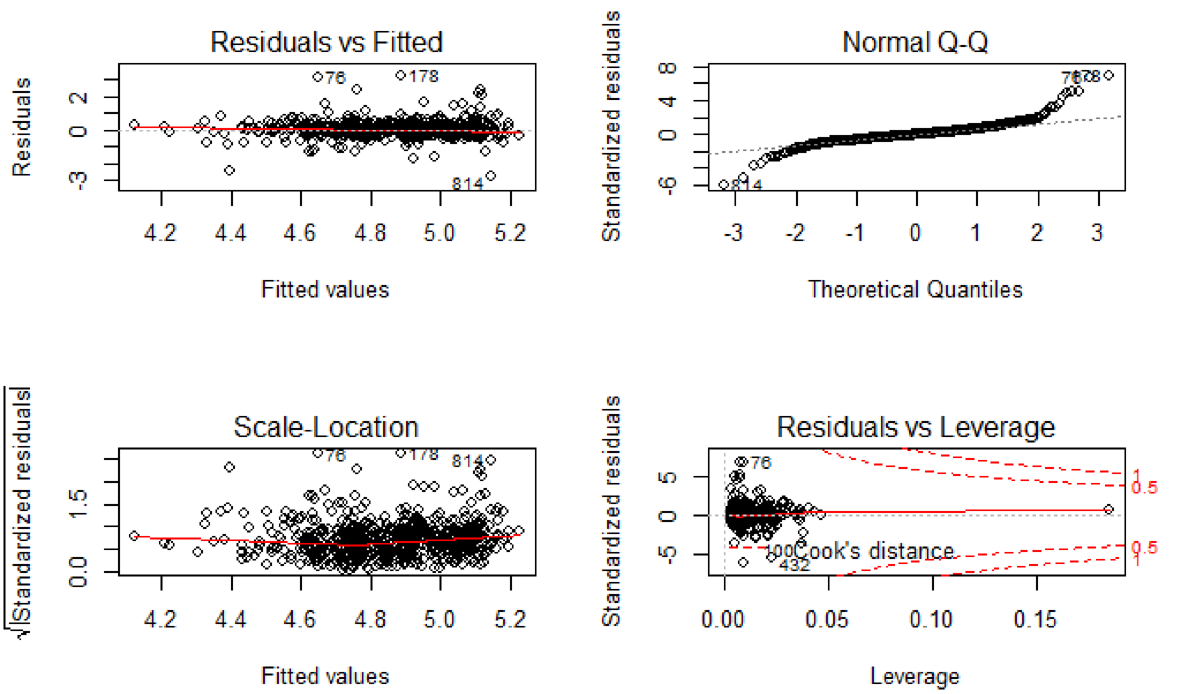
Mincer 1997



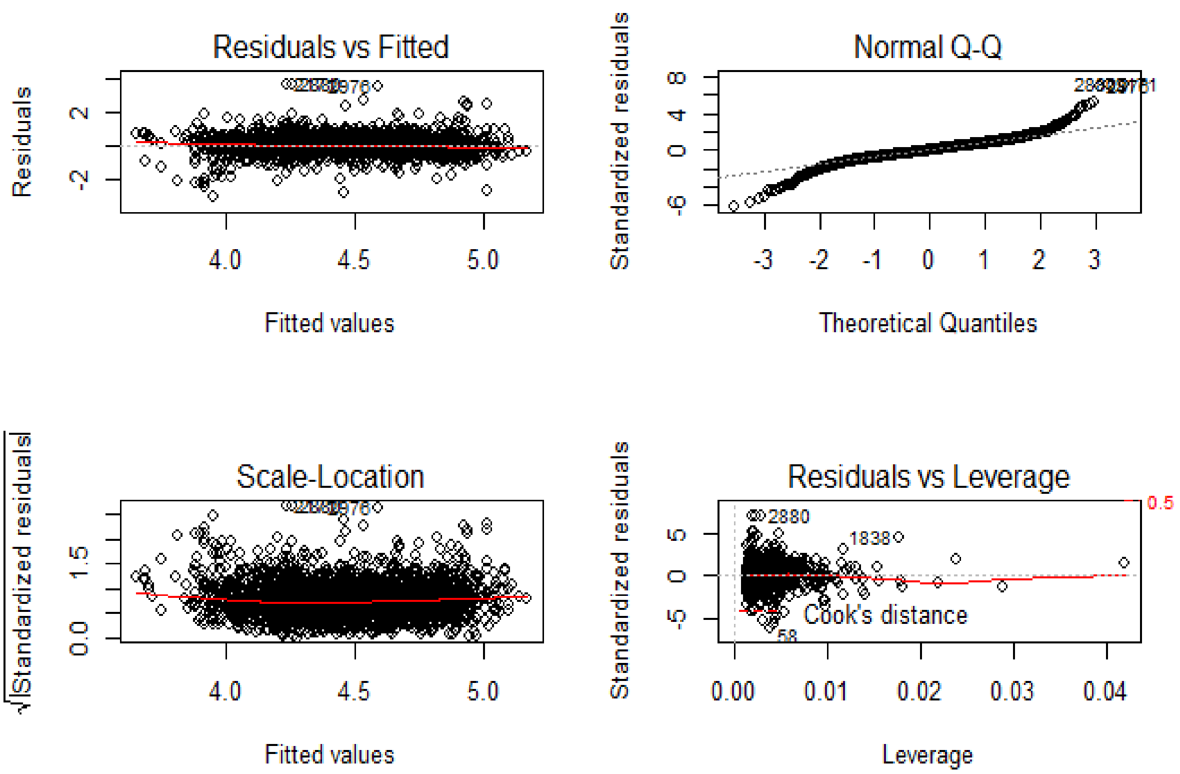
Mincer 2005



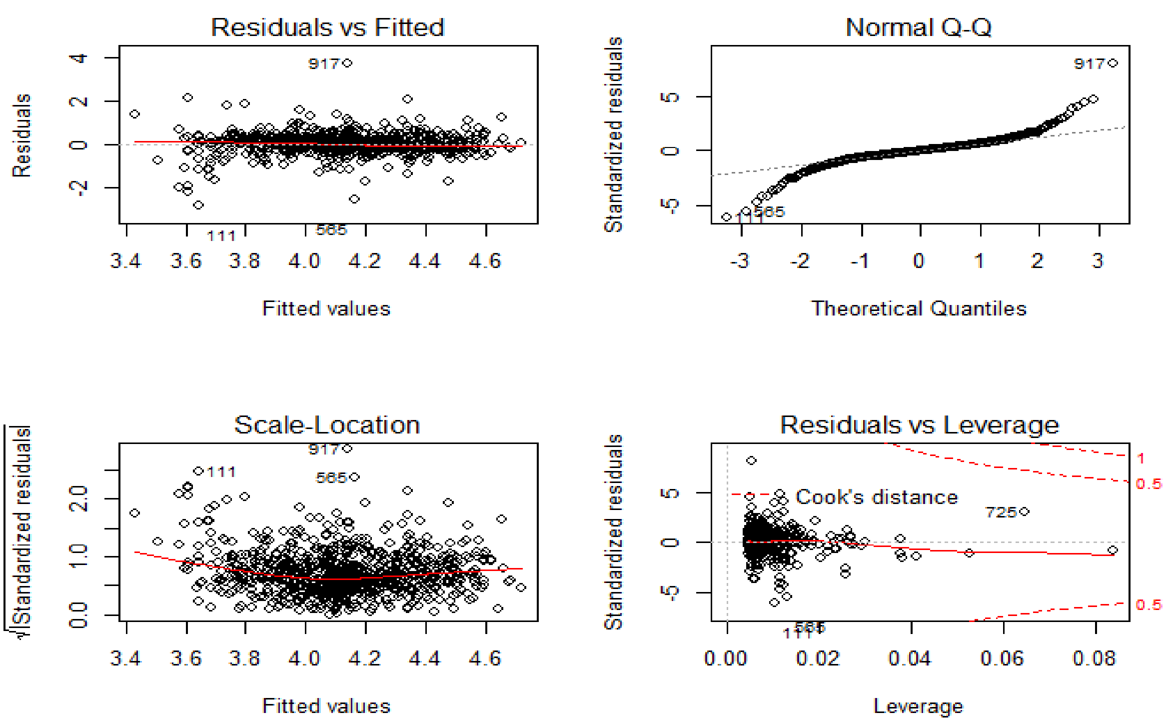
Mincer 2015



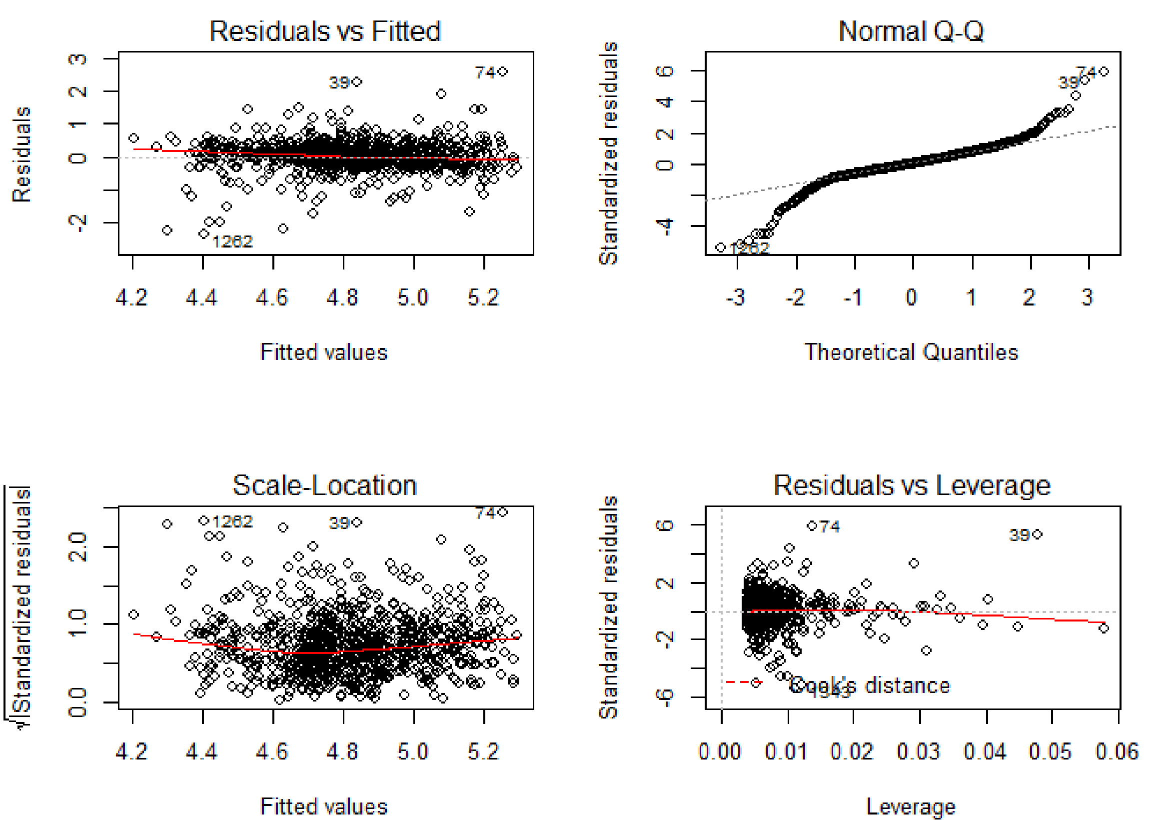
Mincer total



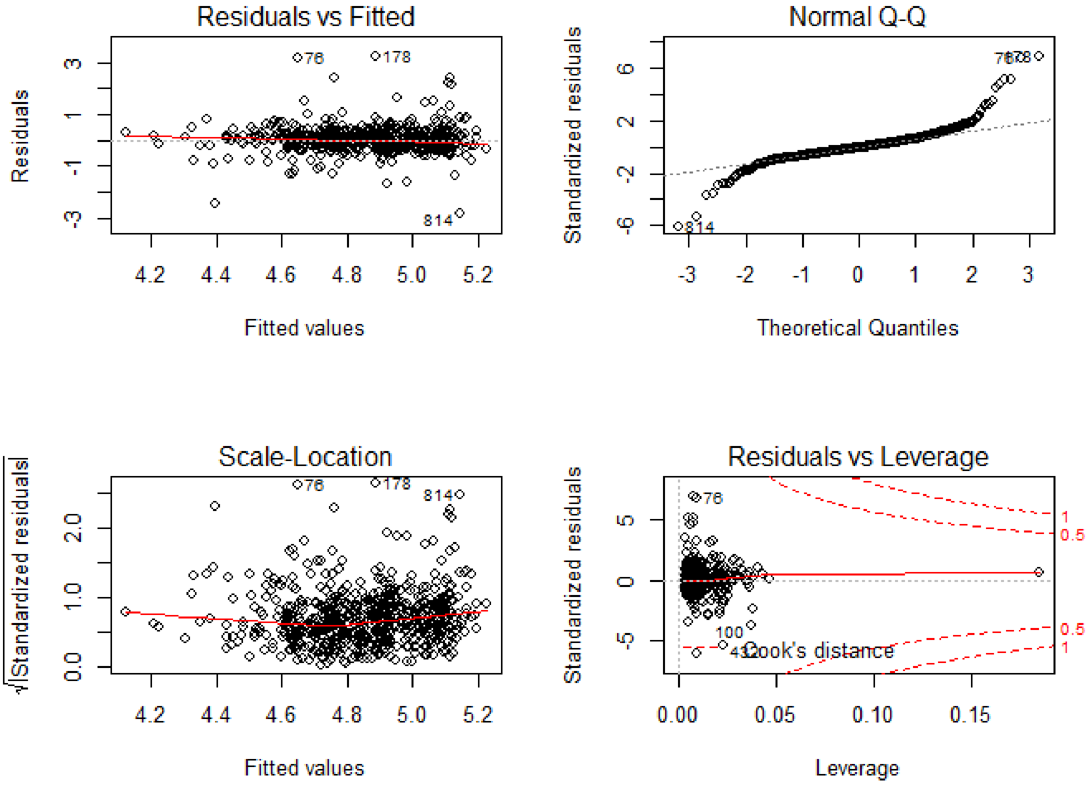
Mincer 1997, degree dummy



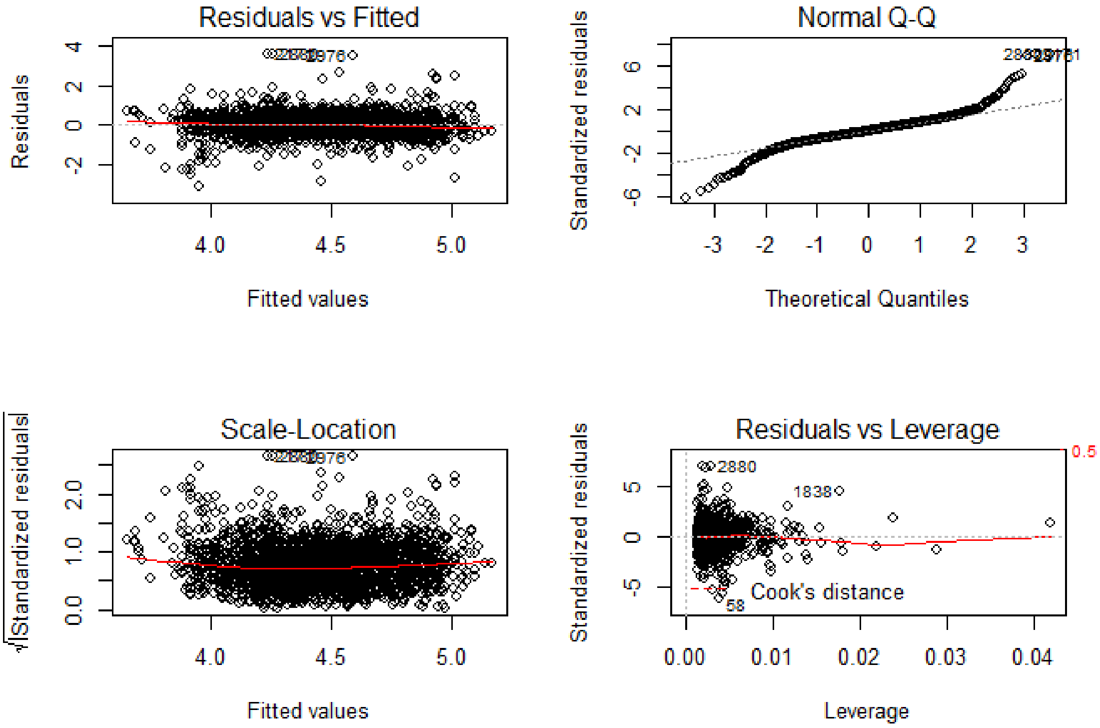
Mincer 2005, degree dummy



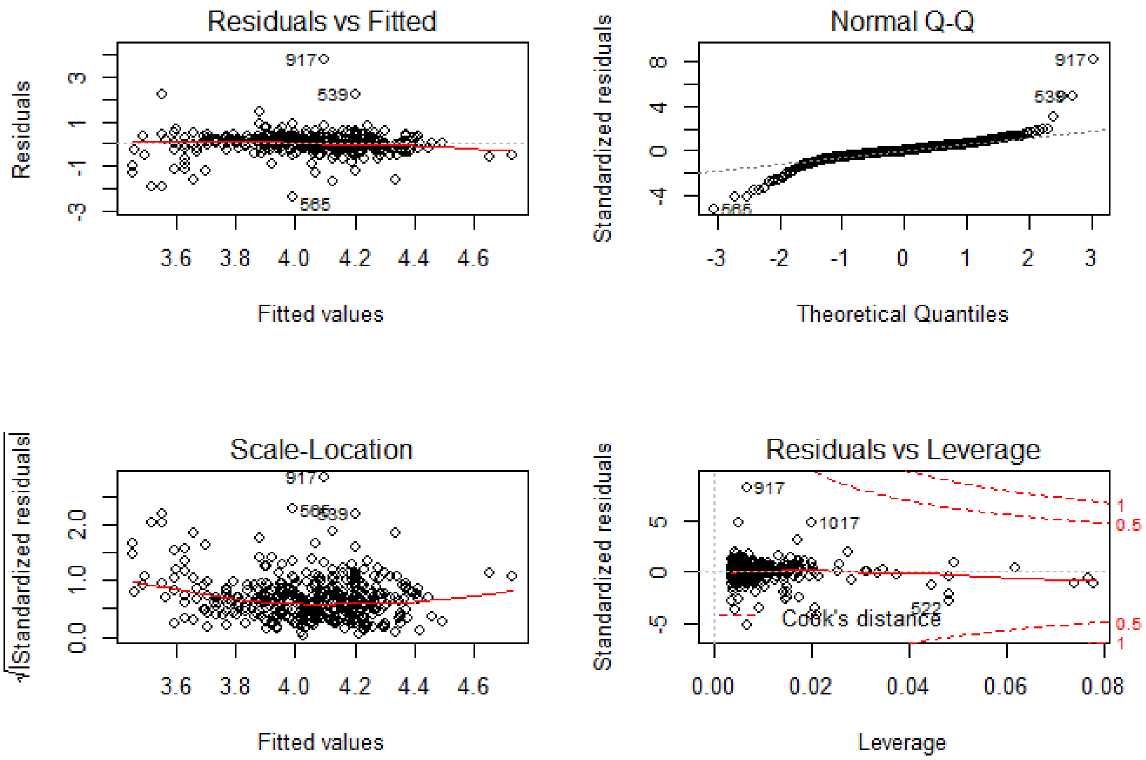
Mincer 2015, degree dummy



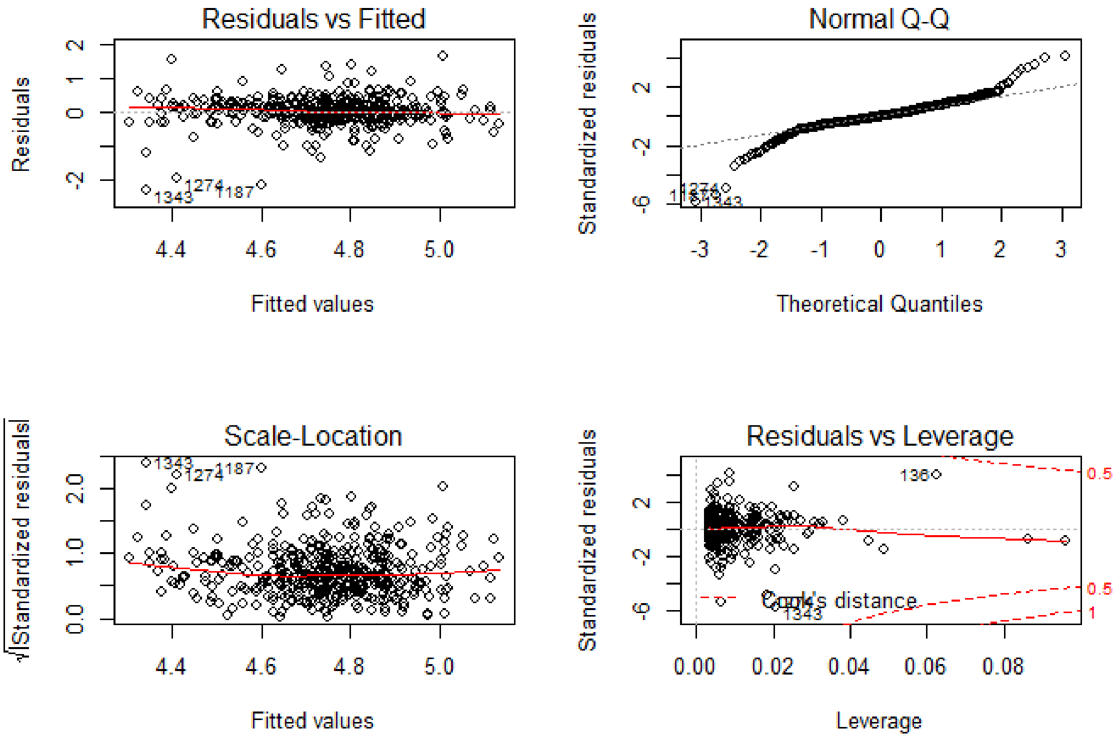
Mincer total dummy degree



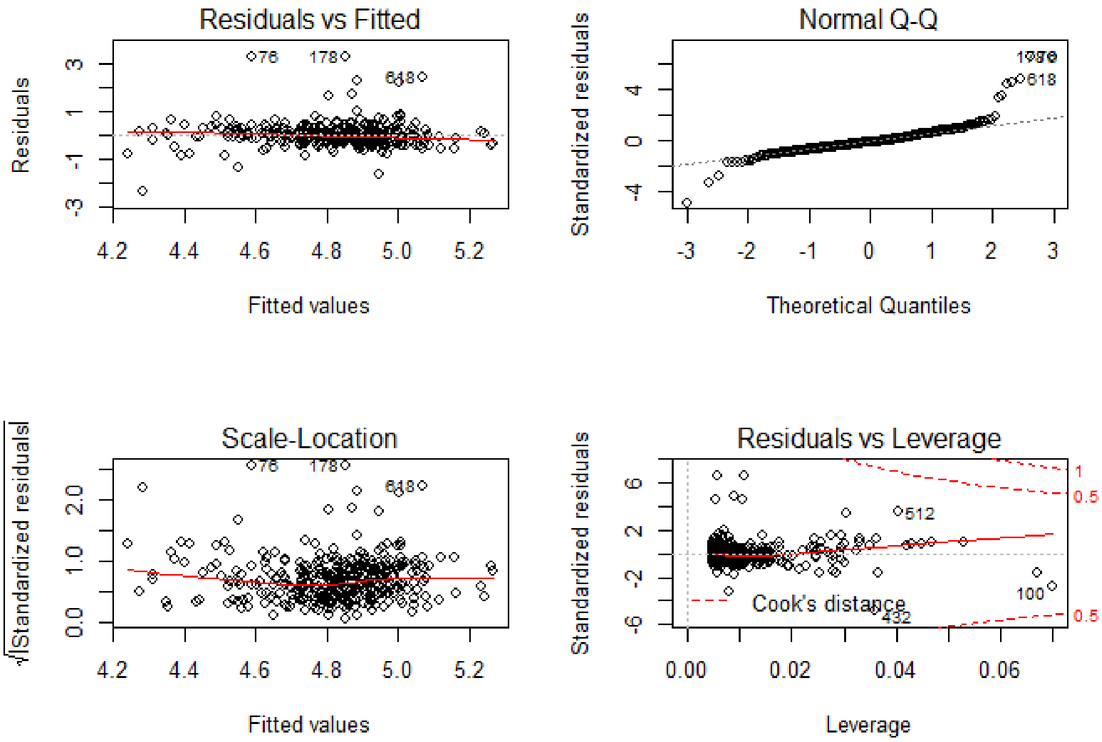
Mincer 1997, females only



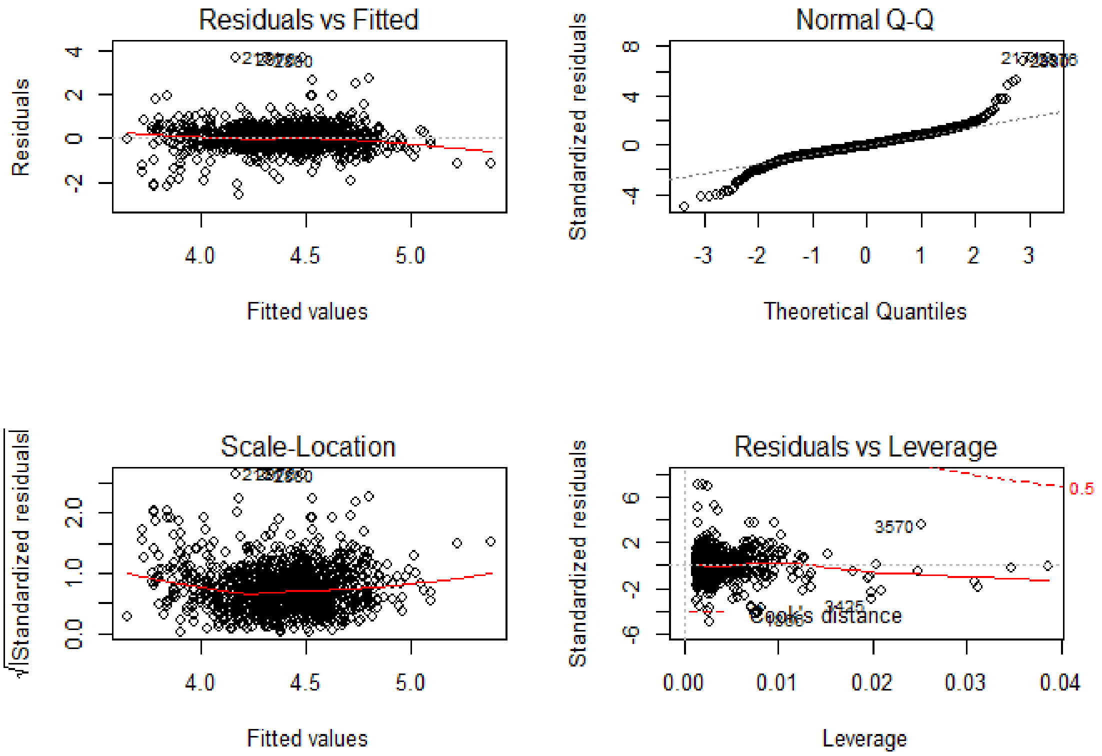
Mincer 2005, females only



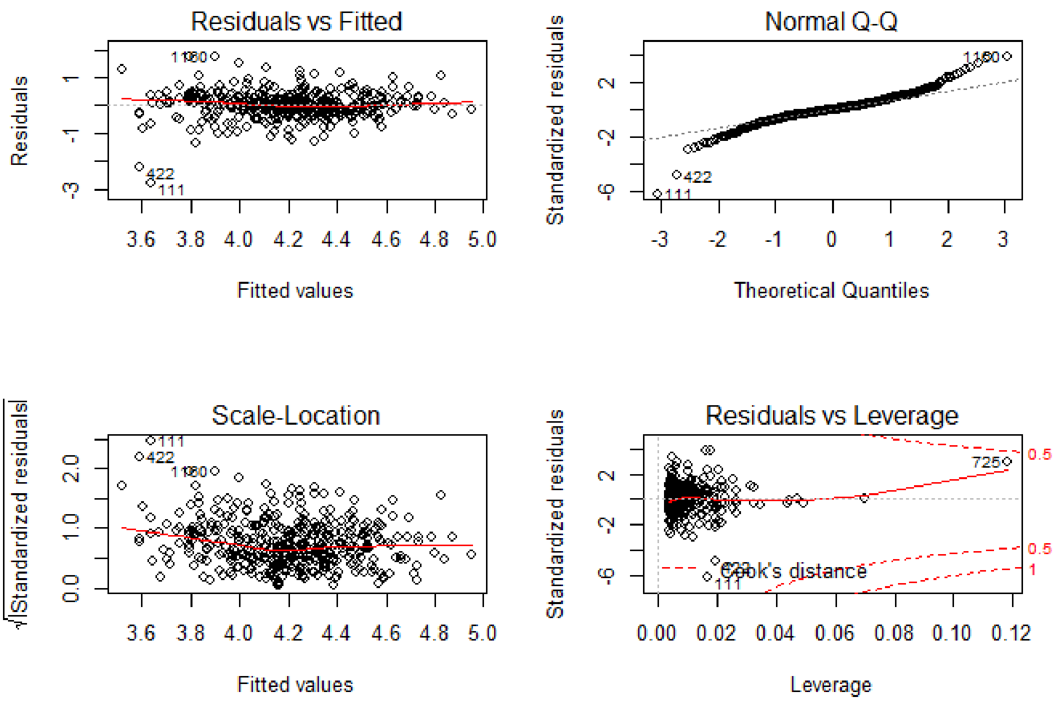
Mincer 2015, females only



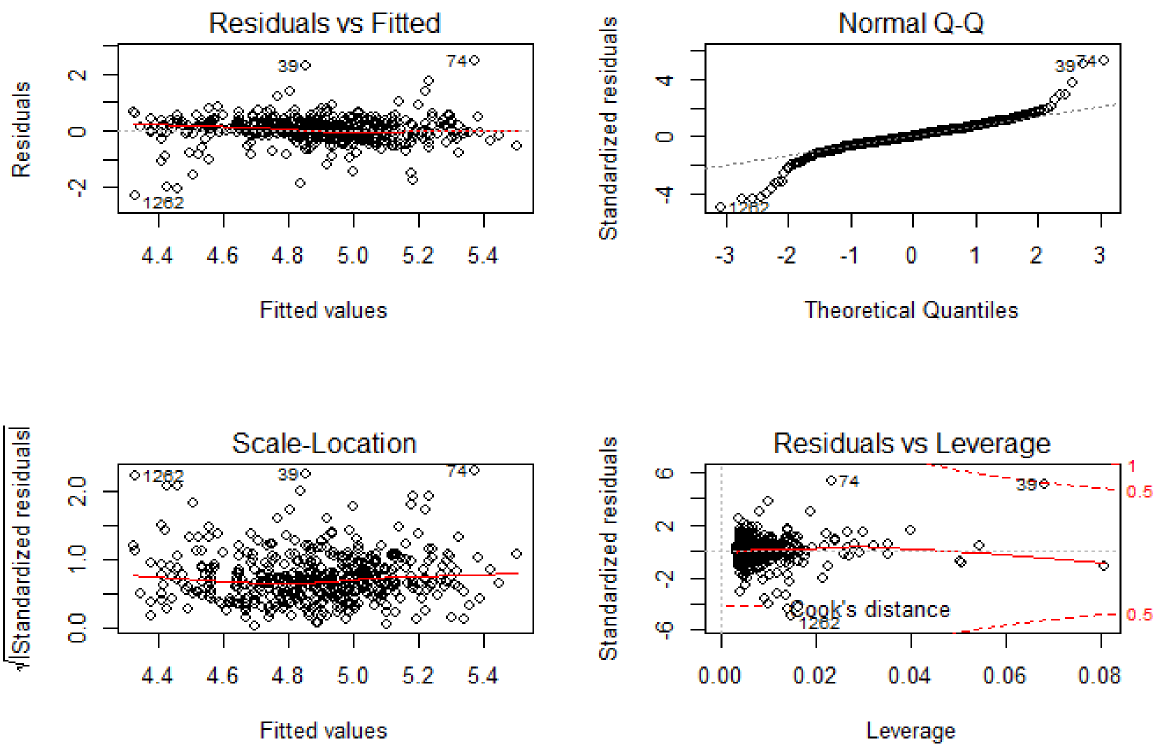
Mincer total, females only



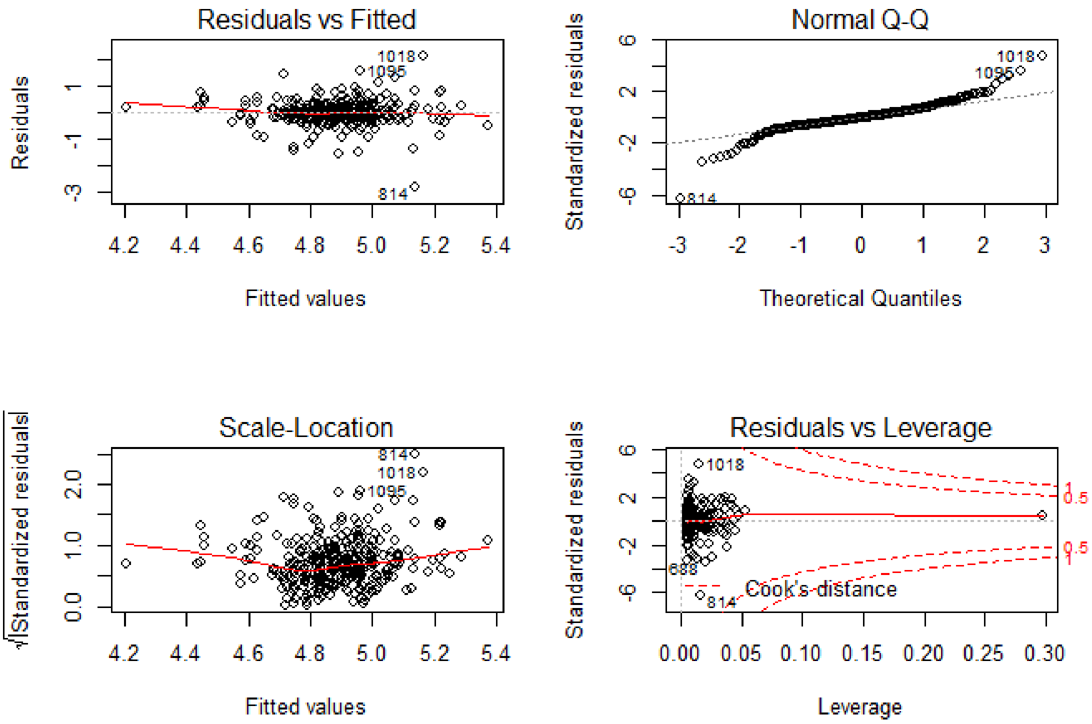
Mincer 1997, males only



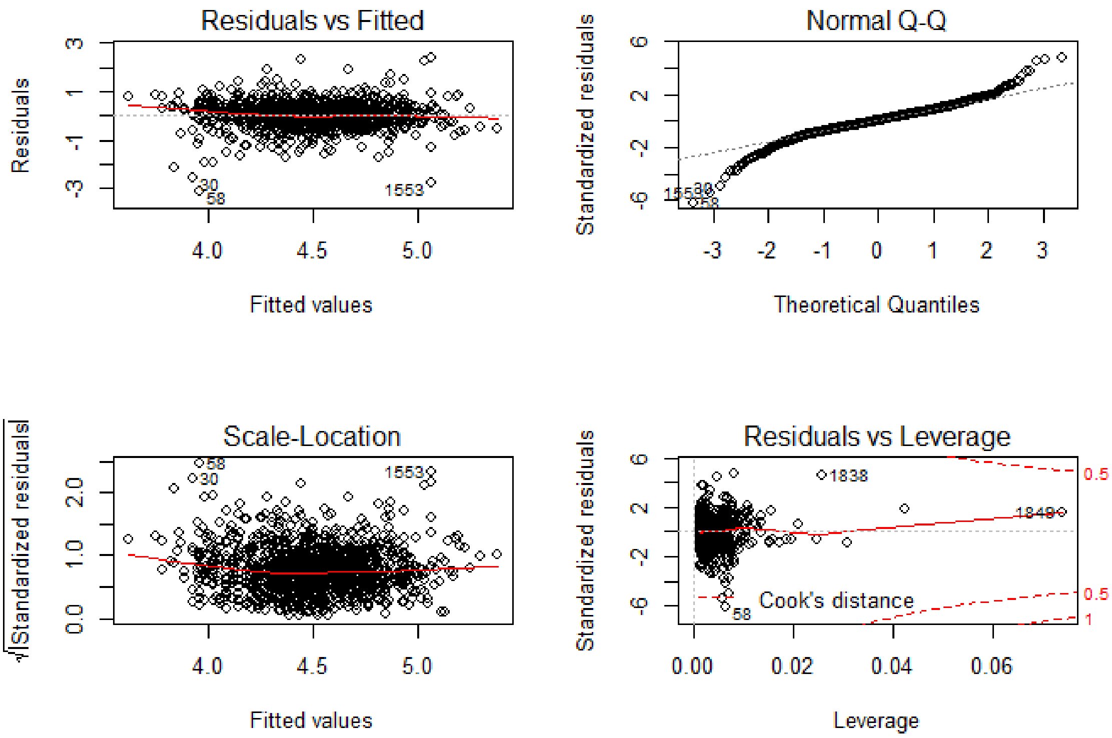
Mincer 2005, males only



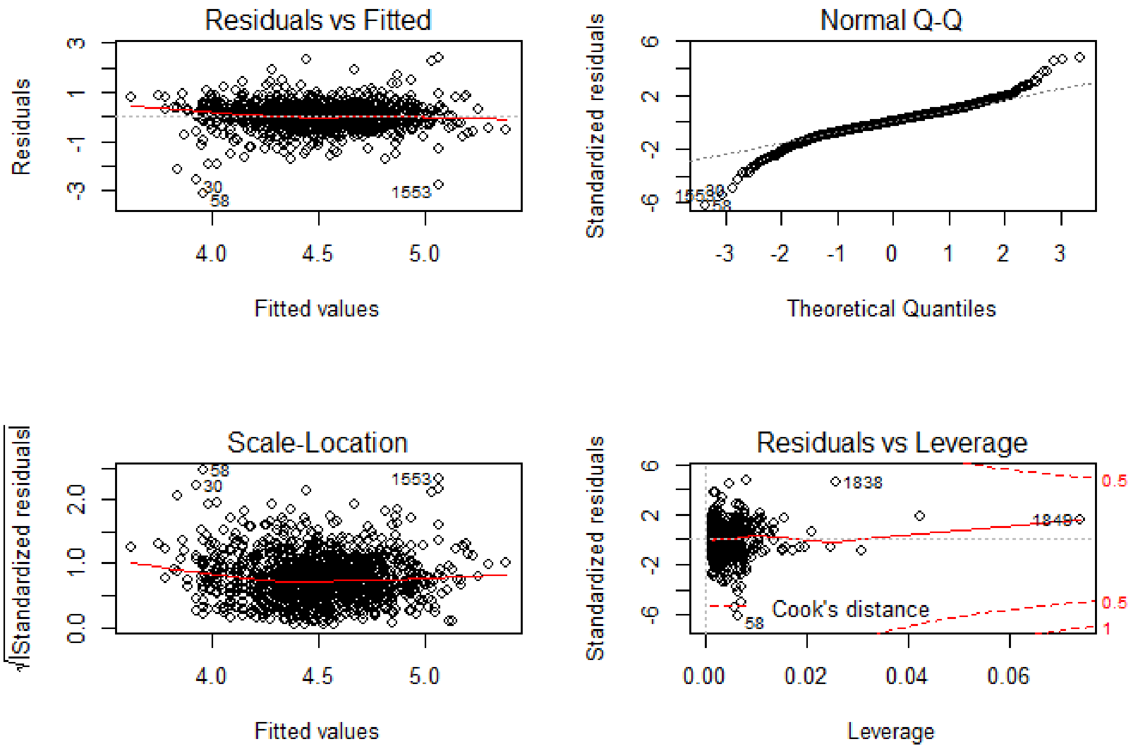
Mincer 2015, males only



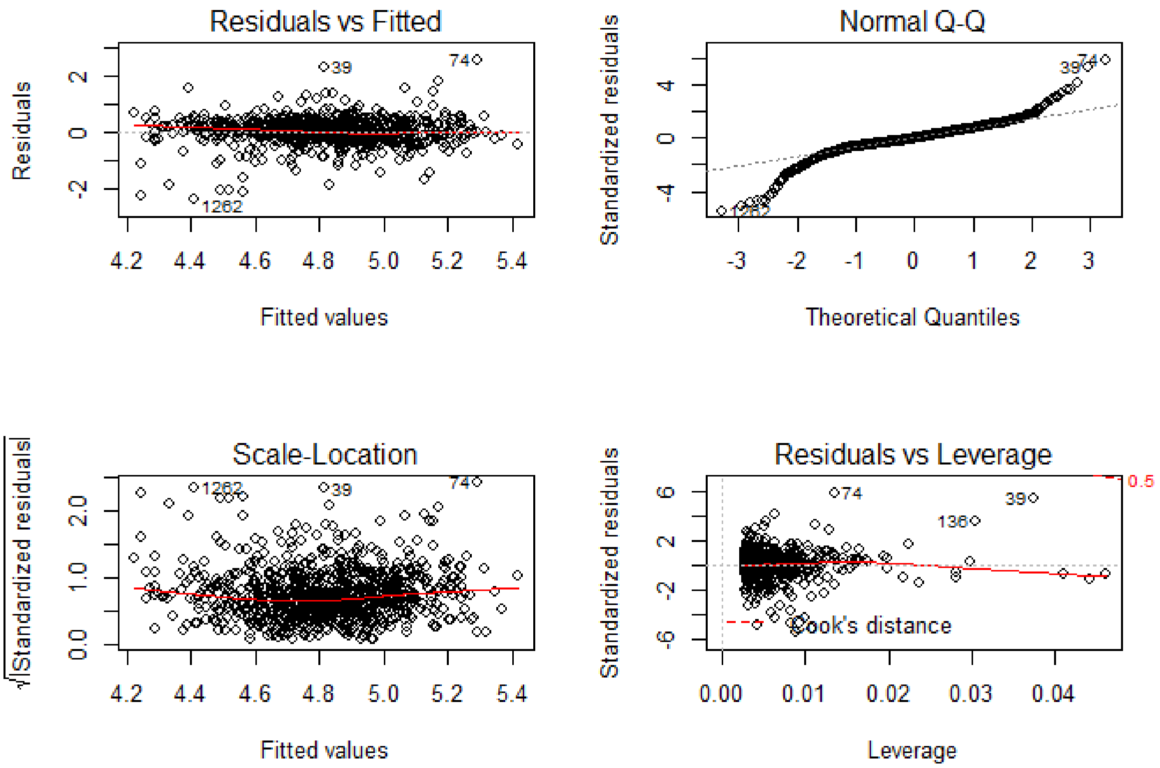
Mincer total males only



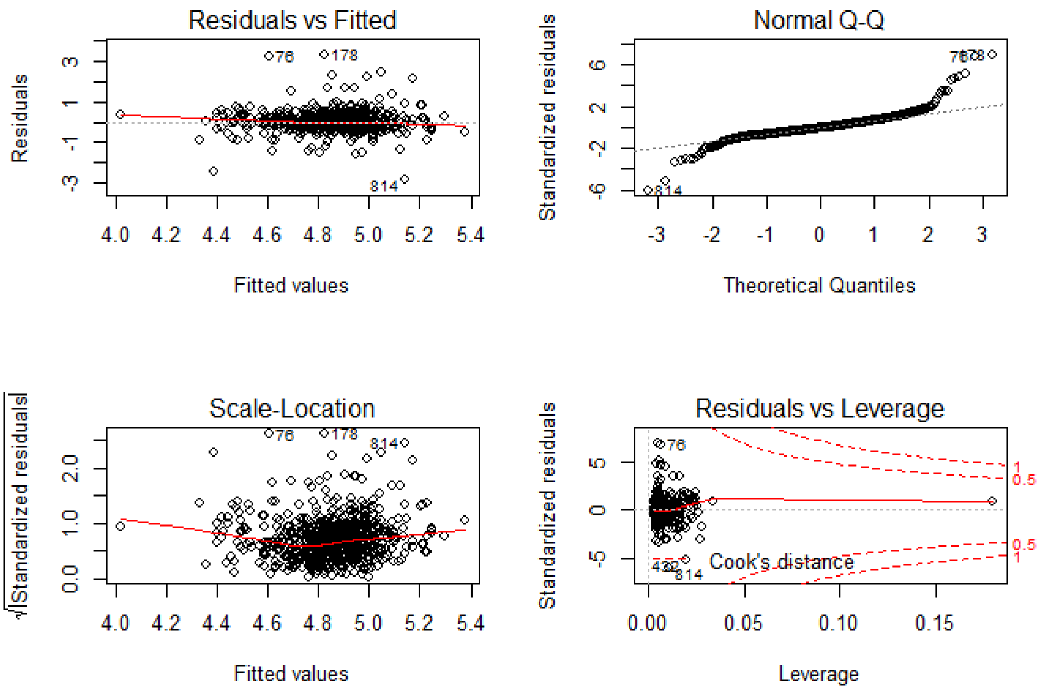
Mincer 1997, dummy sex



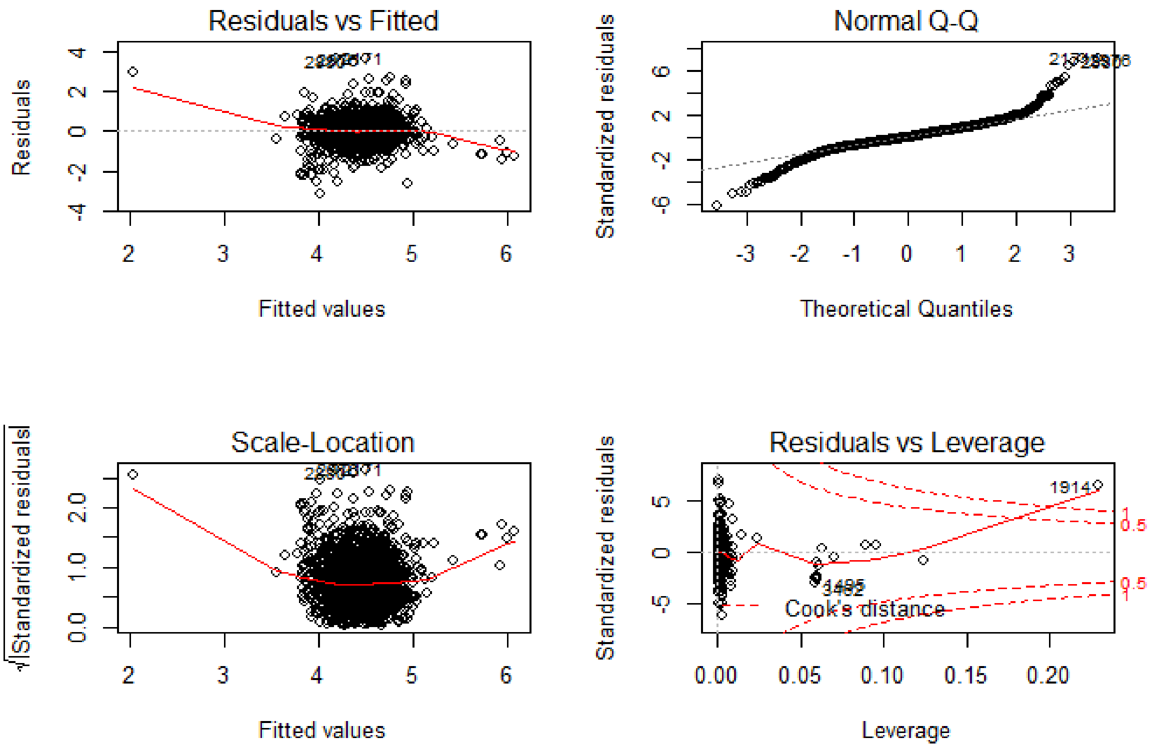
Mincer 2005, Dummy sex



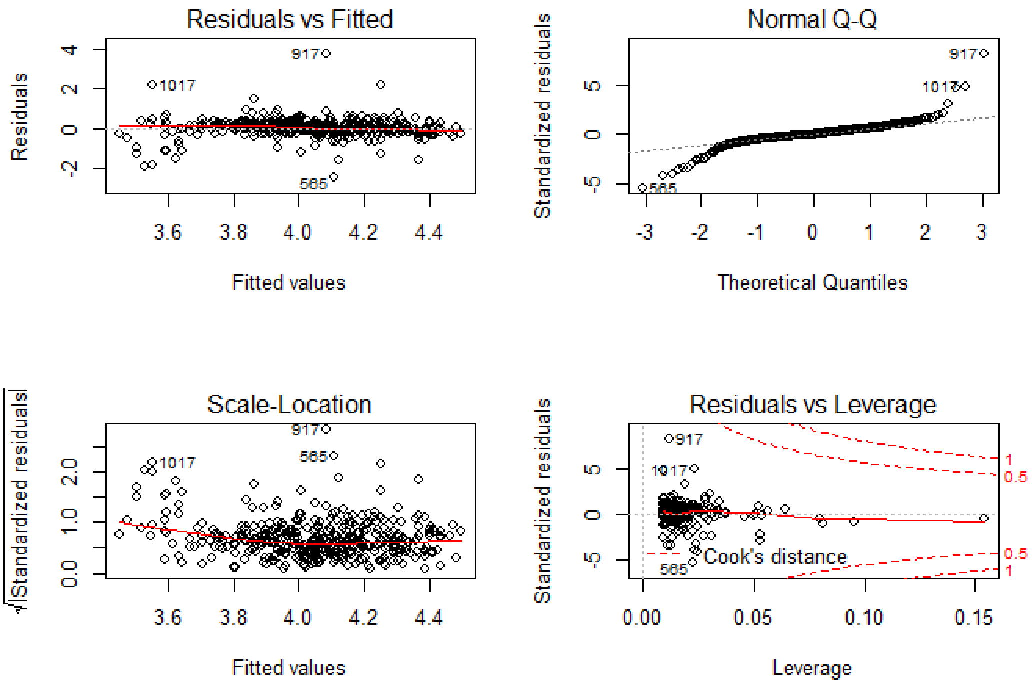
Mincer 2015, dummy sex



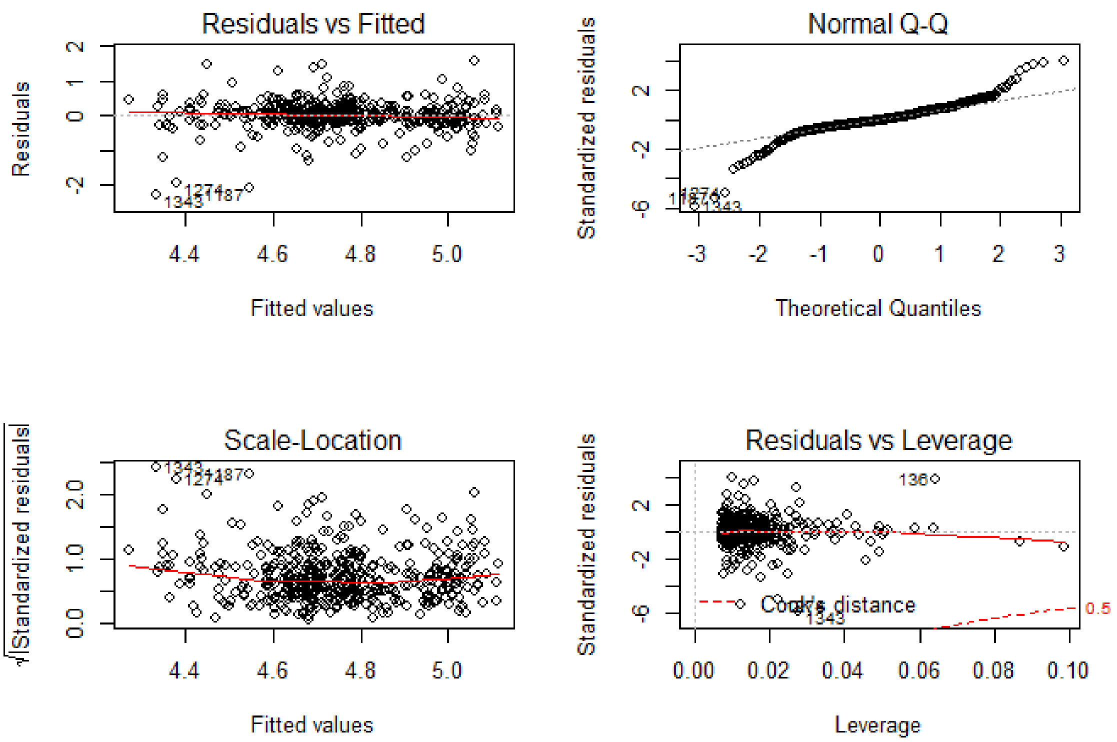
Mincer total, dummy sex



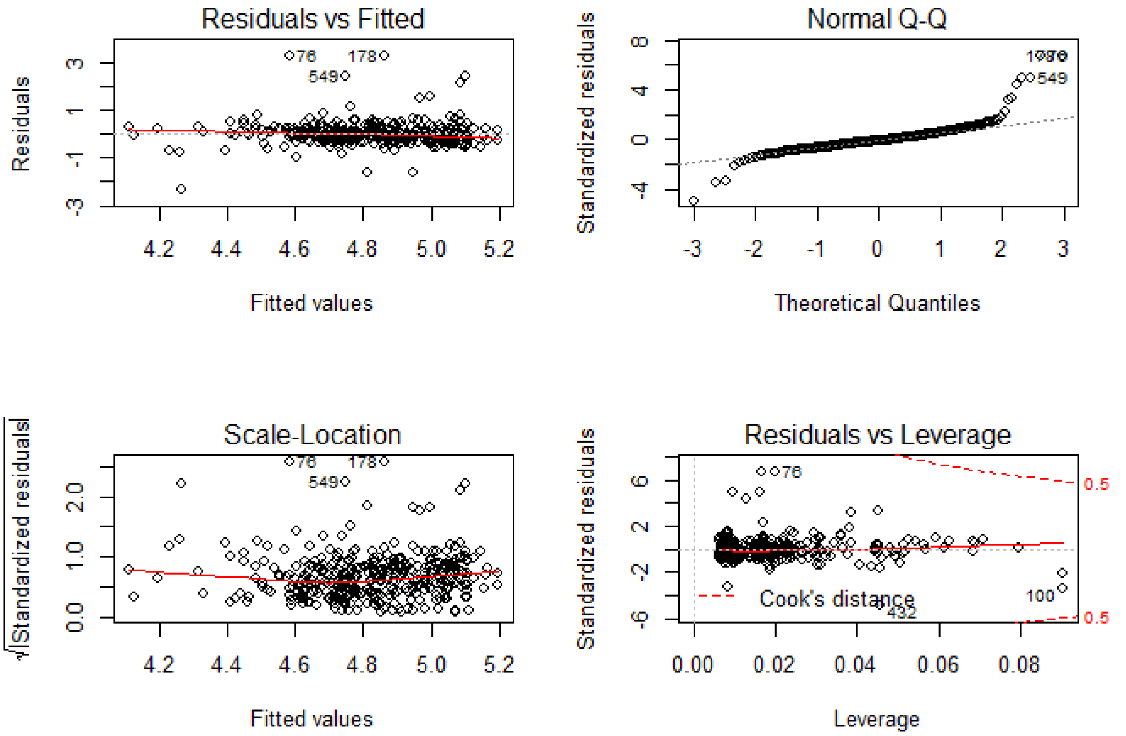
Mincer 1997 (female set), Mincer with dummy for degree



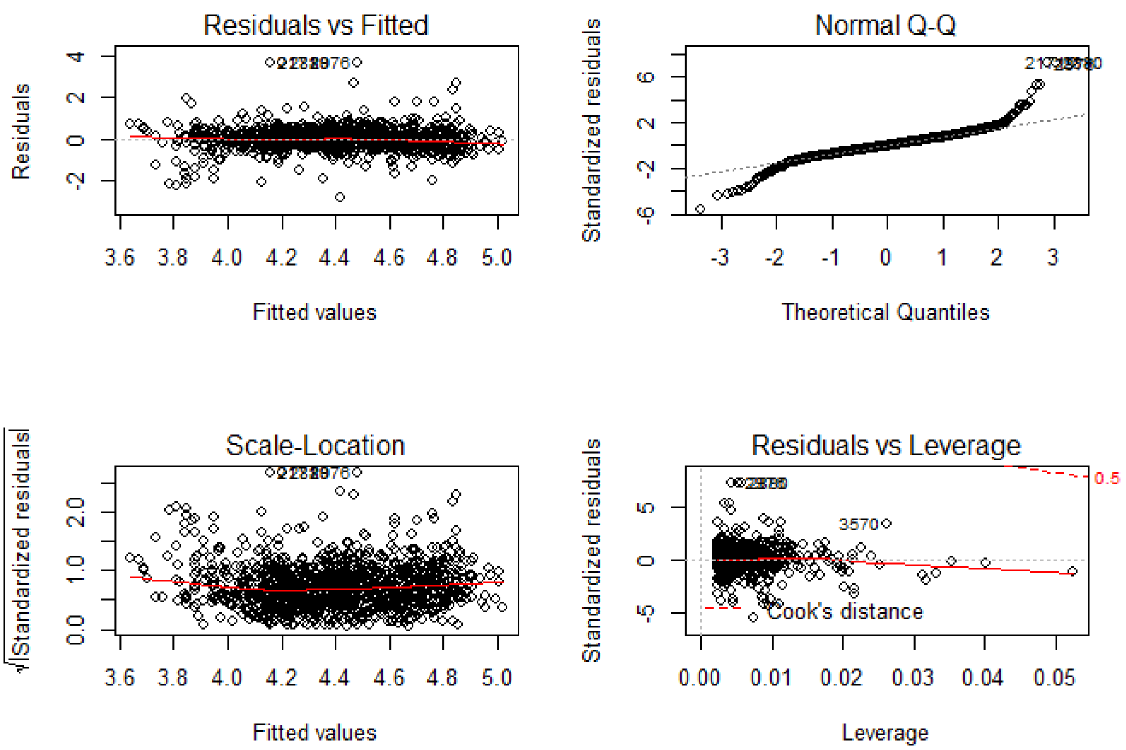
Mincer 2005 (female set), Mincer with dummy for degree



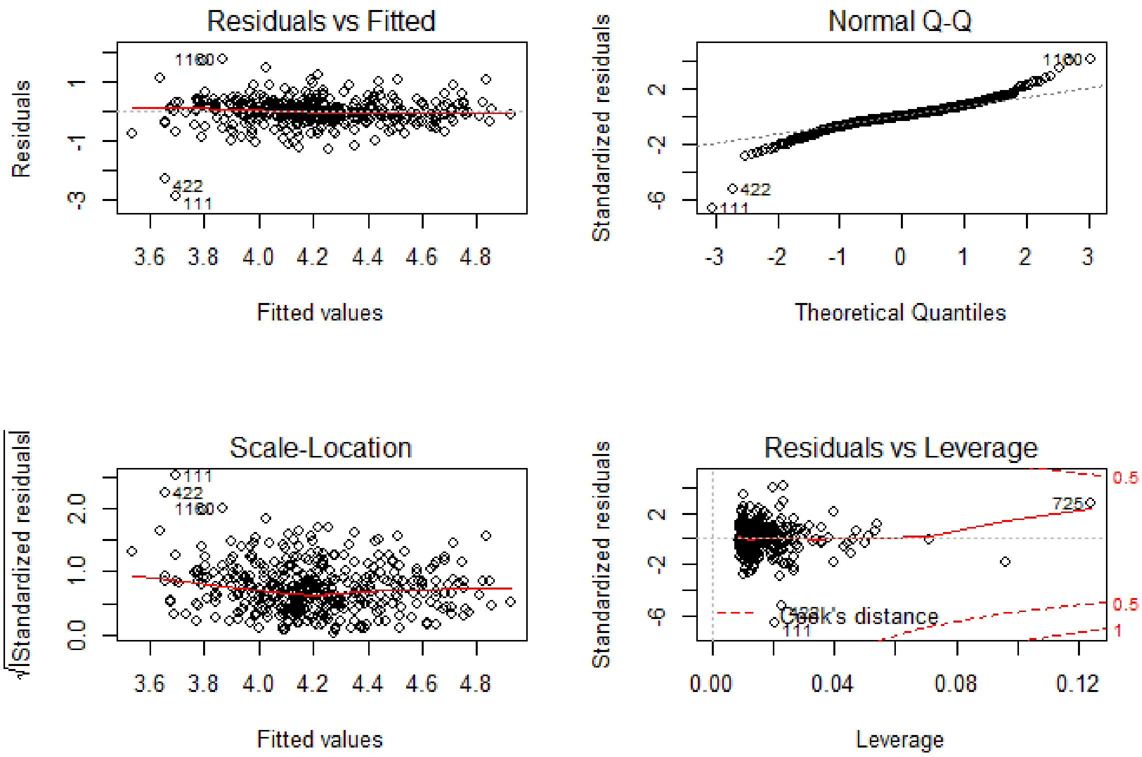
Mincer 2015 (female set), Mincer with dummy for degree



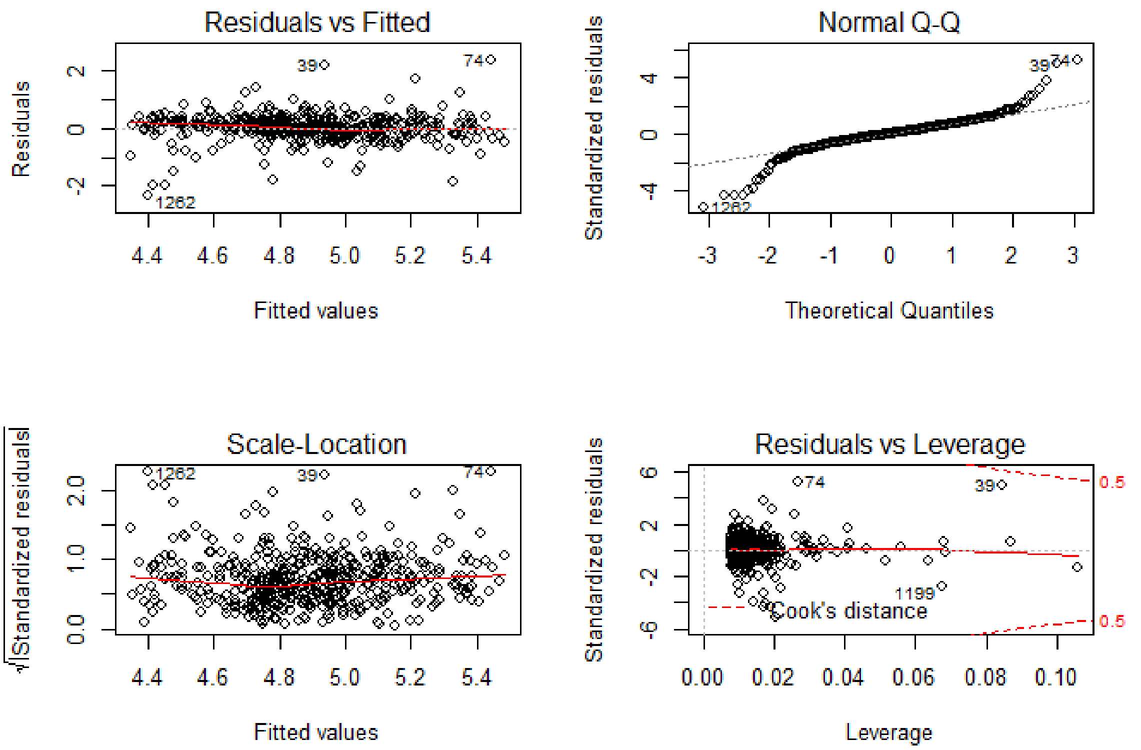
Mincer total (female set), Mincer with dummy for degree



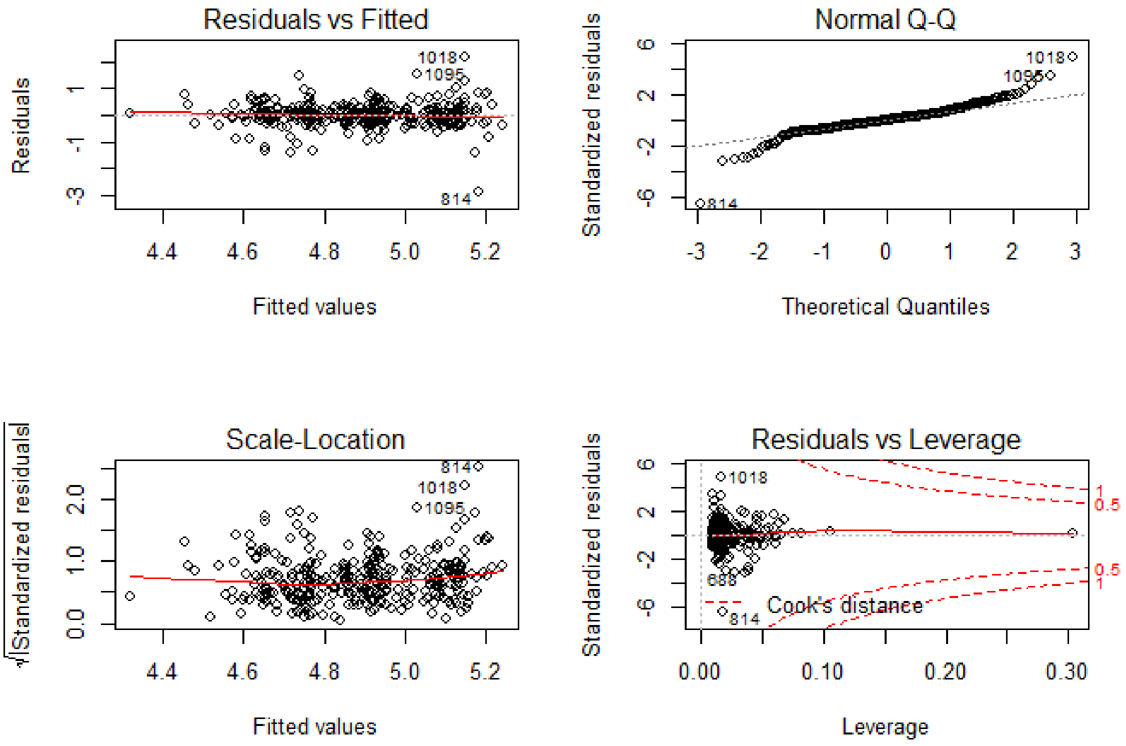
Mincer 1997 (male set), Mincer with dummy for degree



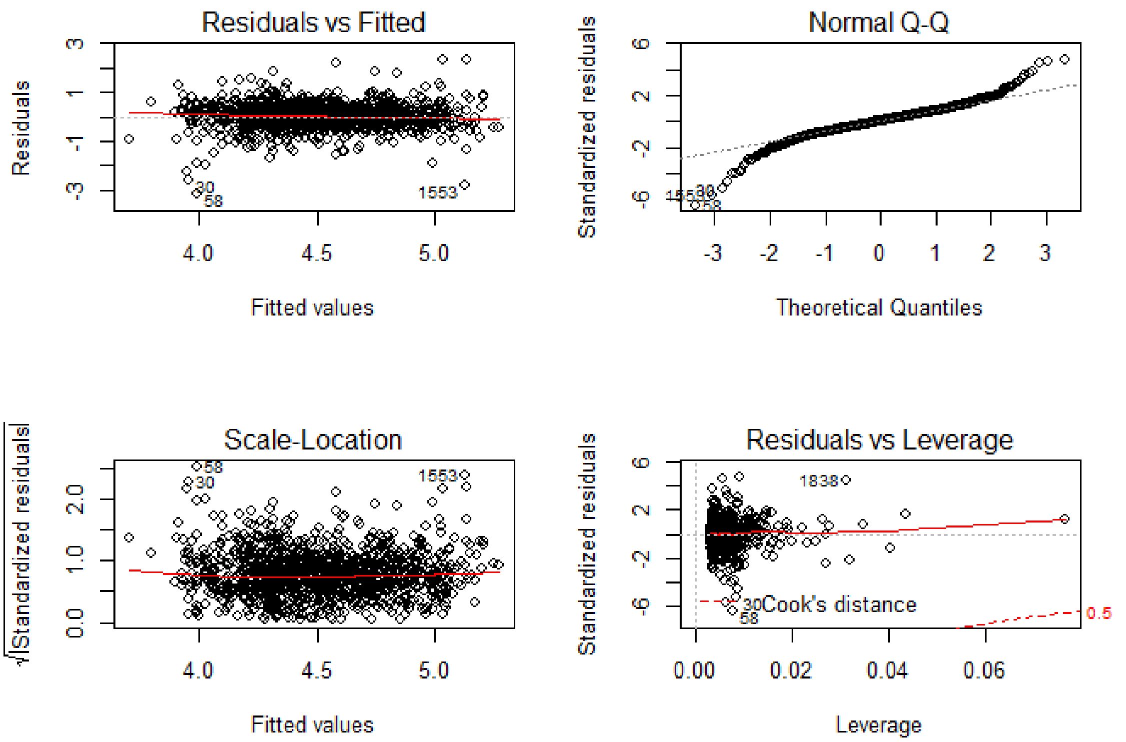
Mincer 2005 (male set), Mincer with dummy for degree



Mincer 2015 (male set), Mincer with dummy for degree



Mincer total (total), Mincer with dummy for degree



Appendix 3: Tables

Table 1.1 1997

group:1997	mean	sd	median	min	max	range	se
Female	0.50	0.50	0.00	0.00	1.00	1.00	0.02
Age	43.52	12.94	44.00	18.00	75.00	57.00	0.42
Schooling	12.16	3.69	12.00	1.00	34.00	33.00	0.12
Hours	38.31	9.36	40.00	1.00	70.00	69.00	0.30
Wage	17025	30943	15000	1000	910000	909000.00	1021
Exp	23.74	13.94	23.00	0.00	66.00	66.00	0.46
Whour	78.66	212.92	61.05	2.29	5833	5831	7.03
lnwage	4.12	0.55	4.11	0.83	8.67	7.84	0.02
D_Vocat	0.23	0.42	0.00	0.00	1.00	1.00	0.01
D_Gym	0.25	0.44	0.00	0.00	1.00	1.00	0.01
D_Uni	0.23	0.42	0.00	0.00	1.00	1.00	0.01
D_Comp	0.29	0.45	0.00	0.00	1.00	1.00	0.01

Table 1.2 2005

group:2005	mean	sd	median	min	max	range	se
Female	0.49	0.50	0.00	0.00	1.00	1.00	0.02
Age	44.02	13.32	44.00	18.00	79.00	61.00	0.42
Schooling	12.49	3.67	12.00	1.00	25.00	24.00	0.12
Hours	37.64	9.80	40.00	2.00	80.00	78.00	0.31
Wage	21383	10052	20000	1000	100000	99000.00	321
Whour	139	111	121.15	7.69	2538	2530	3.55
lnwage	4.81	0.48	4.80	2.04	7.84	5.80	0.02
Exp	24.01	14.15	24.00	0.00	63.00	63.00	0.45
D_Vocat	0.29	0.45	0.00	0.00	1.00	1.00	0.01
D_Gym	0.26	0.44	0.00	0.00	1.00	1.00	0.01
D_Uni	0.27	0.44	0.00	0.00	1.00	1.00	0.01
D_Comp	0.19	0.39	0.00	0.00	1.00	1.00	0.01

Table 1.3 2015

group:2015	mean	sd	median	min	max	range	se
Female	0.53	0.50	1.00	0.00	1.00	1.00	0.02
Age	46.70	12.28	47.00	18.00	79.00	61.00	0.46
Schooling	14.19	3.37	14.00	0.00	24.00	24.00	0.13
Hours	38.64	8.98	40.00	4.00	70.00	66.00	0.33
Wage	35912	46984	30000	1000	840000	839000.00	177
Whour	152	196	121	6.87	3365	3358	7.45
lnwage	4.85	0.50	4.80	1.93	8.12	6.19	0.02
Exp	26.43	13.21	26.50	0.00	73.00	73.00	0.50
D_Vocat	0.21	0.41	0.00	0.00	1.00	1.00	0.02
D_Gym	0.25	0.43	0.00	0.00	1.00	1.00	0.02
D_Uni	0.47	0.50	0.00	0.00	1.00	1.00	0.02
D_Comp	0.07	0.26	0.00	0.00	1.00	1.00	0.01

Table 1.4 Total

group:total	mean	sd	median	min	max	range	se
Female	0.50	0.50	1.00	0.00	1.00	1.00	0.01
Age	44.55	12.98	45.00	18.00	79.00	61.00	0.25
Schooling	12.84	3.69	13.00	0.00	34.00	34.00	0.07
Hours	38.14	9.44	40.00	1.00	80.00	79.00	0.18
Wage	23744	32049	20000	1000	910000	909000.00	629
Whour	105	171	84.13	2.29	5833	5831	3.37
Exp	25.30	13.84	25.00	0.00	73.00	73.00	0.27
lnwage	4.44	0.58	4.43	0.83	8.67	7.84	0.01
D_Vocat	0.25	0.43	0.00	0.00	1.00	1.00	0.01
D_Gym	0.25	0.43	0.00	0.00	1.00	1.00	0.01
D_Uni	0.31	0.46	0.00	0.00	1.00	1.00	0.01
D_Comp	0.19	0.39	0.00	0.00	1.00	1.00	0.01

Table 1.5 Overview of observations

Year:	Start	After removing NA	After removing 0 wrkhrs	After removing both
1997	1275	1177	972	874
2005	1371	1290	1028	947
2015	1162	1122	724	684
Total	3808	3602	2724	2518

Table 2.1 1997, Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.06073	2.900 – 3.222	<.001
Schooling	0.04268	0.033 – 0.053	<.001
Exp	0.04141	0.034 – 0.049	<.001
Exp2	-0.00058	-0.001 – -0.000	<.001
Observations	874		
R^2 / adj. R^2	.1892 / .1864		

Table 2.2 2005, Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.89099	3.744 – 4.038	<.001
Schooling	0.04077	0.032 – 0.050	<.001
Exp	0.02835	0.021 – 0.035	<.001
Exp2	-0.00034	-0.000 – -0.000	<.001
Observations	947		
R^2 / adj. R^2	.1521 / .1494		

Table 2.3 2015, Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.92851	3.709 – 4.149	<.001
Schooling	0.03968	0.028 – 0.051	<.001
Exp	0.02472	0.015 – 0.034	<.001
Exp2	-0.00033	-0.001 – -0.000	<.001
Observations	684		
R^2 / adj. R^2	.0949 / .0909		

Table 2.4 Total, Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.30583	3.205 – 3.407	<.001
Schooling	0.04808	0.044 – 0.053	<.001
Exp	0.03401	0.031 – 0.037	<.001
Exp2	-0.00041	-0.000 – -0.000	<.001
Observations	2518		
R^2 / adj. R^2	.1690 / .1680		

Table 3.1 1997, Mincer with dummy for degree

	Lnwage		
	B	*CI*	*p*
(Intercept)	3.15988	2.974 – 3.345	<.001
Schooling	0.02999	0.017 – 0.043	<.001
Exp	0.03996	0.032 – 0.048	<.001
Exp2	-0.00056	-0.001 – -0.000	<.001
D_Vocat	0.01497	-0.081 – 0.111	.760
D_Gym	0.08679	-0.016 – 0.189	.097
D_Uni	0.18429	0.065 – 0.303	.002
Observations	859		
R^2 / adj. R^2	.1986 / .1930		

Table 3.2 2005, Mincer with dummy for degree

	Lnwage		
	B	*CI*	*p*
(Intercept)	3.94438	3.773 – 4.116	<.001
Schooling	0.02415	0.013 – 0.036	<.001
Exp	0.02869	0.022 – 0.036	<.001
Exp2	-0.00034	-0.000 – -0.000	<.001
D_Vocat	0.09423	-0.001 – 0.190	.053
D_Gym	0.16296	0.058 – 0.268	.002
D_Uni	0.27091	0.155 – 0.387	<.001
Observations	942		
R^2 / adj. R^2	.1724 / .1671		

Table 3.3 2015, Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	4.01773	3.779 – 4.257	<.001
Schooling	0.01556	0.001 – 0.030	.036
Exp	0.02557	0.016 – 0.035	<.001
Exp2	-0.00033	-0.001 – -0.000	<.001
D_Vocat	0.07497	-0.085 – 0.235	.359
D_Gym	0.20693	0.044 – 0.370	.013
D_Uni	0.35992	0.186 – 0.534	<.001
Observations	683		
R^2 / adj. R^2	.1331 / .1254		

Table 3.4 total, Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.22231	3.102 – 3.343	<.001
Schooling	0.03543	0.027 – 0.044	<.001
Exp	0.03395	0.029 – 0.039	<.001
Exp2	-0.00038	-0.000 – -0.000	<.001
D_Vocat	0.16281	0.096 – 0.230	<.001
D_Gym	0.23324	0.161 – 0.305	<.001
D_Uni	0.39377	0.313 – 0.474	<.001
Observations	2484		
R^2 / adj. R^2	.2190 / .2171		

Table 4.1 1997, Mincer with dummy for sex

	lnwage		
	B	*CI*	*p*
(Intercept)	3.12405	2.963 – 3.285	<.001
Schooling	0.04429	0.035 – 0.054	<.001
Exp	0.04111	0.034 – 0.049	<.001
Exp2	-0.00058	-0.001 – -0.000	<.001
Female	-0.15653	-0.218 – -0.095	<.001
Observations	874		
R^2 / adj. R^2	.2118 / .2082		

Table 4.2 2005, Mincer with dummy for sex

	lnwage		
	B	*CI*	*p*
(Intercept)	3.94668	3.800 – 4.093	<.001
Schooling	0.04189	0.033 – 0.051	<.001
Exp	0.02933	0.023 – 0.036	<.001
Exp2	-0.00036	-0.000 – -0.000	<.001
Female	-0.16042	-0.216 – -0.105	<.001
Observations	947		
R^2 / adj. R^2	.1798 / .1764		
Observations	2430		
R^2 / adj. R^2	.2532 / .2476		

Table 4.3 2015, Mincer with dummy for sex

	lnwage		
	B	*CI*	*p*
(Intercept)	3.96058	3.739 – 4.182	<.001
Schooling	0.04045	0.029 – 0.052	<.001
Exp	0.02459	0.015 – 0.034	<.001
Exp2	-0.00033	-0.001 – -0.000	<.001
Female	-0.07243	-0.145 – 0.000	.050
Observations	684		
R^2 / adj. R^2	.1000 / .0947		

Table 4.4 Total, Mincer with dummy for sex

	lnwage		
	B	*CI*	*p*
(Intercept)	3.35757	3.255 – 3.460	<.001
Schooling	0.04869	0.044 – 0.053	<.001
Exp	0.03431	0.031 – 0.038	<.001
Exp2	-0.00042	-0.000 – -0.000	<.001
Female	-0.11913	-0.160 – -0.078	<.001
Observations	2518		
R^2 / adj. R^2	.1797 / .1784		

Table 5.1 1997 (female dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.09502	2.862 – 3.328	<.001
Schooling	0.03483	0.020 – 0.049	<.001
Exp	0.04189	0.031 – 0.053	<.001
Exp2	-0.00063	-0.001 – -0.000	<.001
Observations	432		
R^2 / adj. R^2	.1675 / .1616		

Table 5.2 1997 (male dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	2.92106	2.694 – 3.148	<.001
Schooling	0.05255	0.039 – 0.066	<.001
Exp	0.04348	0.033 – 0.054	<.001
Exp2	-0.00054	-0.001 – -0.000	<.001
Observations	442		
R^2 / adj. R^2	.2338 / .2285		

Table 5.3 2005 (female dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.96821	3.771 – 4.165	<.001
Schooling	0.03378	0.022 – 0.045	<.001
Exp	0.02517	0.016 – 0.034	<.001
Exp2	-0.00033	-0.001 – -0.000	<.001
Observations	467		
R^2 / adj. R^2	.1264 / .1207		

Table 5.4 2005 (male dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.79039	3.578 – 4.002	<.001
Schooling	0.04889	0.036 – 0.062	<.001
Exp	0.03329	0.023 – 0.043	<.001
Exp2	-0.00039	-0.001 – -0.000	<.001
Observations	480		
R^2 / adj. R^2	.1954 / .1903		

Table 5.5 2015 (female dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.77655	3.445 – 4.108	<.001
Schooling	0.03907	0.022 – 0.056	<.001
Exp	0.03351	0.019 – 0.048	<.001
Exp2	-0.00049	-0.001 – -0.000	.002
Observations	361		
R^2 / adj. R^2	.1070 / .0995		

Table 5.6 2015 (male dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	4.03166	3.728 – 4.335	<.001
Schooling	0.04302	0.027 – 0.059	<.001
Exp	0.01499	0.002 – 0.028	.028
Exp2	-0.00018	-0.000 – 0.000	.149
Observations	323		
R^2 / adj. R^2	.0953 / .0868		

Table 5.7 Total (female dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.10994	2.952 – 3.268	<.001
Schooling	0.05799	0.049 – 0.067	<.001
Exp	0.03576	0.028 – 0.043	<.001
Exp2	-0.00046	-0.001 – -0.000	<.001
Observations	1260		
R^2 / adj. R^2	.1762 / .1742		

Table 5.8 Total (male dataset), Mincer

	lnwage		
	B	*CI*	*p*
(Intercept)	3.12331	2.975 – 3.272	<.001
Schooling	0.06347	0.055 – 0.072	<.001
Exp	0.03568	0.028 – 0.043	<.001
Exp2	-0.00040	-0.001 – -0.000	<.001
Observations	1245		
R^2 / adj. R^2	.2174 / .2155		

Table 6.1 1997 (female dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.18099	2.911 – 3.451	<.001
Schooling	0.02434	0.006 – 0.042	.008
Exp	0.04224	0.030 – 0.054	<.001
Exp2	-0.00065	-0.001 – -0.000	<.001
D_Vocat	-0.00054	-0.140 – 0.139	.994
D_Gym	0.03663	-0.110 – 0.184	.625
D_Uni	0.13168	-0.025 – 0.288	.099
Observations		426	
R^2 / adj. R^2		.1758 / .1640	

Table 6.2 1997 (male dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.08981	2.830 – 3.349	<.001
Schooling	0.03411	0.016 – 0.053	<.001
Exp	0.04019	0.030 – 0.051	<.001
Exp2	-0.00049	-0.001 – -0.000	<.001
D_Vocat	0.02132	-0.108 – 0.150	.745
D_Gym	0.11682	-0.024 – 0.258	.104
D_Uni	0.26937	0.089 – 0.449	.003
Observations		433	
R^2 / adj. R^2		.2526 / .2421	

Table 6.3 2005 (female dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	4.10341	3.875 – 4.332	<.001
Schooling	0.01578	0.001 – 0.030	.034
Exp	0.02405	0.015 – 0.033	<.001
Exp2	-0.00032	-0.001 – -0.000	<.001
D_Vocat	0.05402	-0.068 – 0.176	.386
D_Gym	0.07034	-0.070 – 0.210	.324
D_Uni	0.24760	0.107 – 0.388	<.001
Observations	464		
R^2 / adj. R^2	.1580 / .1470		

Table 6.4 2005 (male dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.84389	3.598 – 4.090	<.001
Schooling	0.02882	0.012 – 0.046	.001
Exp	0.03466	0.024 – 0.045	<.001
Exp2	-0.00041	-0.001 – -0.000	<.001
D_Vocat	0.12039	-0.022 – 0.263	.097
D_Gym	0.19658	0.044 – 0.349	.012
D_Uni	0.34257	0.159 – 0.527	<.001
Observations	478		
R^2 / adj. R^2	.2182 / .2083		

Table 6.5 2015 (female dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.83622	3.480 – 4.193	<.001
Schooling	0.01888	-0.001 – 0.039	.066
Exp	0.03276	0.018 – 0.047	<.001
Exp2	-0.00046	-0.001 – -0.000	.003
D_Vocat	0.06324	-0.173 – 0.300	.600
D_Gym	0.18544	-0.057 – 0.428	.134
D_Uni	0.33203	0.086 – 0.578	.008
Observations	361		
R^2 / adj. R^2	.1415 / .1269		

Table 6.6 2015 (male dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	4.16969	3.837 – 4.502	<.001
Schooling	0.01294	-0.008 – 0.034	.236
Exp	0.01761	0.004 – 0.031	.011
Exp2	-0.00022	-0.000 – 0.000	.083
D_Vocat	0.08665	-0.132 – 0.305	.435
D_Gym	0.21524	-0.006 – 0.437	.057
D_Uni	0.40696	0.158 – 0.656	.001
Observations	322		
R^2 / adj. R^2	.1415 / .1252		

Table 6.7 total (female dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.22786	3.053 – 3.402	<.001
Schooling	0.03414	0.023 – 0.045	<.001
Exp	0.03380	0.026 – 0.041	<.001
Exp2	-0.00041	-0.001 – -0.000	<.001
D_Vocat	0.14741	0.051 – 0.244	.003
D_Gym	0.17656	0.072 – 0.281	<.001
D_Uni	0.37238	0.264 – 0.480	<.001
Observations	1251		
R^2 / adj. R^2	.2074 / .2036		

Table 6.8 total (male dataset), Mincer with dummy for degree

	lnwage		
	B	*CI*	*p*
(Intercept)	3.23732	3.072 – 3.403	<.001
Schooling	0.03574	0.024 – 0.047	<.001
Exp	0.03482	0.028 – 0.042	<.001
Exp2	-0.00037	-0.001 – -0.000	<.001
D_Vocat	0.17143	0.079 – 0.264	<.001
D_Gym	0.25911	0.160 – 0.359	<.001
D_Uni	0.44352	0.323 – 0.564	<.001
Observations	1233		
R^2 / adj. R^2	.2497 / .2461		

Table 7.1, Mincer with control variables

	lnwage		
	B	*CI*	*p*
(Intercept)	3.38983	3.260 – 3.520	<.001
Schooling	0.03149	0.024 – 0.039	<.001
Exp	0.02950	0.024 – 0.035	<.001
Exp2	-0.00034	-0.000 – -0.000	<.001
Sex	-0.14280	-0.182 – -0.103	<.001
D_Vocat	0.15285	0.087 – 0.219	<.001
D_Gym	0.21960	0.149 – 0.290	<.001
D_Uni	0.40274	0.324 – 0.482	<.001
D_mar	0.09964	0.051 – 0.148	<.001
D_sep	0.07791	-0.009 – 0.165	.080
D_wid	0.18154	-0.024 – 0.387	.083
D_sub	-0.04655	-0.096 – 0.003	.064
D_urb	-0.09102	-0.153 – -0.029	.004
D_smagot	-0.06196	-0.139 – 0.015	.113
D_south	0.02496	-0.039 – 0.089	.444
D_west	-0.00386	-0.063 – 0.055	.898
D_northmid	-0.14590	-0.247 – -0.045	.005
D_north	-0.10018	-0.181 – -0.019	.015
D_stock	0.15968	0.095 – 0.225	<.001
Observations	2430		
R^2 / adj. R^2	.2532 / .2476		

Appendix 4: Figures

Figure 1.1 Total, NA

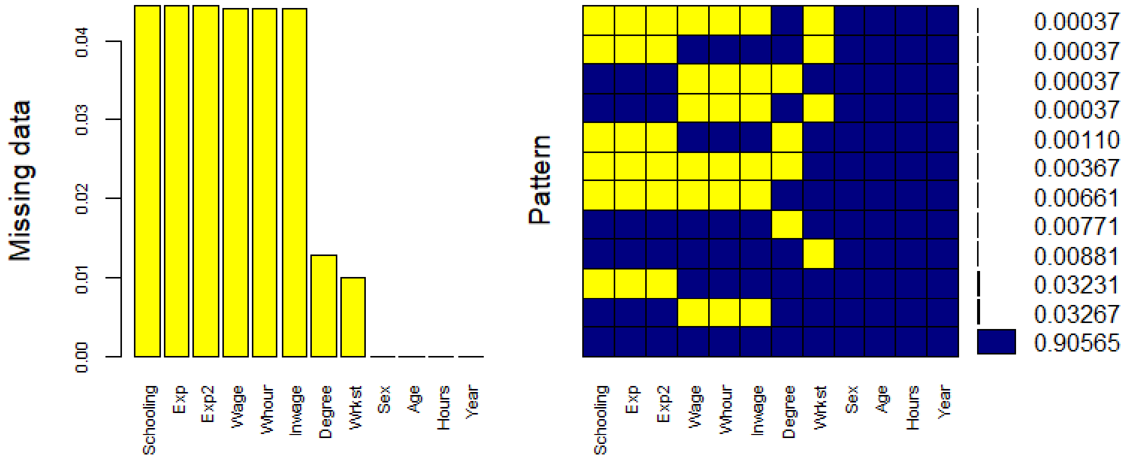
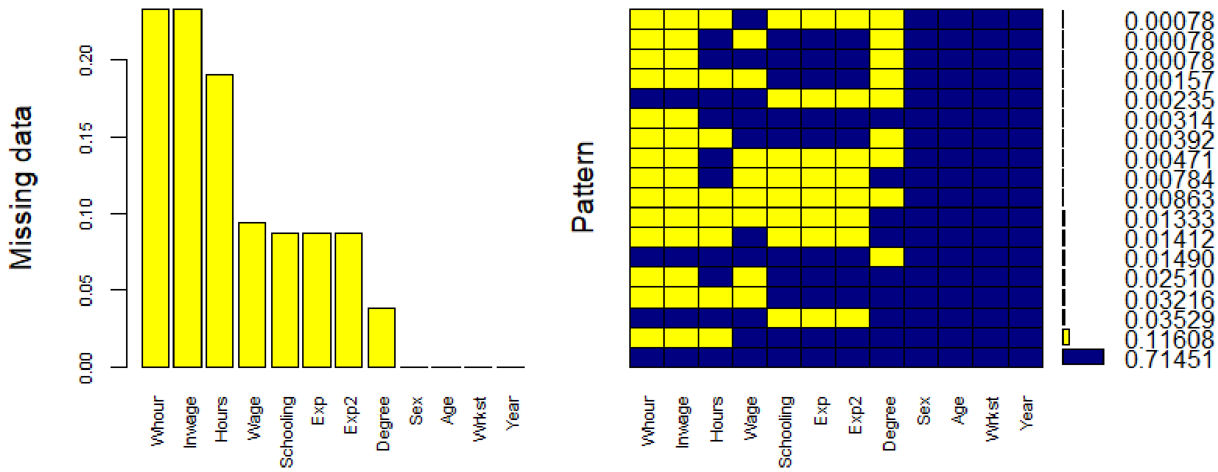


Figure 1.2 1997, NA



Figur 1.3 2005, NA

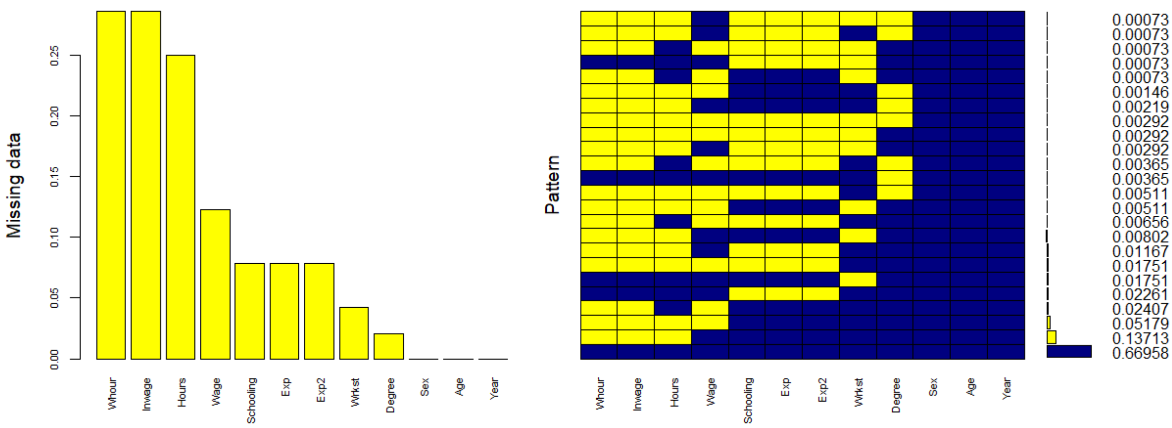


Figure 1.4 2015, NA

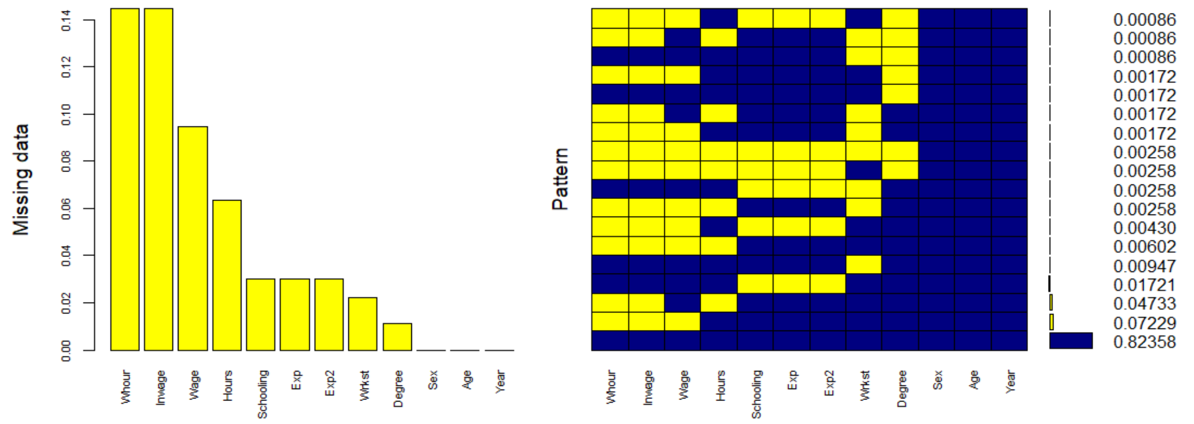


Figure 2.1 histograms work hours

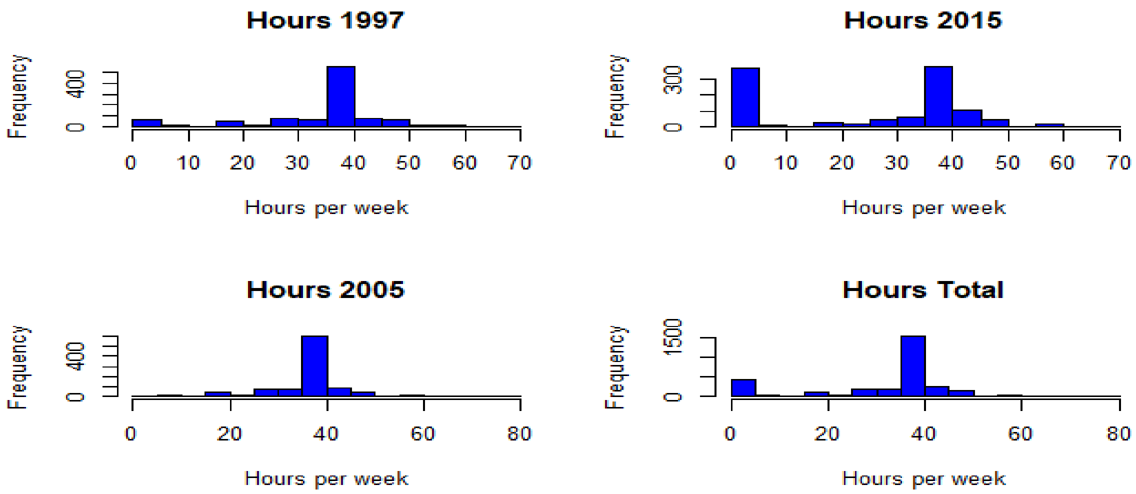


Figure 3.1