How does parental education influence earnings?

Gimy Balaj 14474433
Luca Crijns 14435489
Jędrzej Rękawek 14537591

Course: Introduction Econometrics and Actuarial Science

Lecturer: Rob van Hemert

Group: WG03

Date: 24 March 2023

Contents

Introduction

Human capital is one of the best indicators of someone's success on the job market. People with higher education tend to have higher earnings after graduation. The main purpose of this research project is to investigate if the inheritance of human capital occurs in society. If the answer is affirmative, more resources should be spent on education since it will benefit society in later generations. Moreover, inequalities will emerge because of this mechanism and some taxes might be proposed.

Education is a component of human capital as stated by Gary S. Becker the author of Human Capital (Becker, 1964, p. 17), and as stated before, an increase in human capital is directly correlated to an increase in an individual's earnings. Thus, the purpose of this research is to investigate whether a parent's education shows correlation to the earnings of their child. Therefore, the research question is: How does the education of parents influence the descendant's earnings? David Card (2005) showed in his work that parental education is indeed correlated with the child's education, but there was no evidence that parental education is correlated with one's earnings.

To investigate the topic the research group will carry out empirical data research. The research group would have to clean the dataset first i.e., removing the outliers and the data points which are not possible. Afterwards multiple linear regression will be carried out, using wage as the dependent variable and parental education together with the interaction and control variables as independent variables. The results of this regression will be used to answer the research question.

Firstly, the sufficient economic theory on human capital and education will be used to provide a prediction to an answer to the research question which can be found in Theoretical framework. How the data will be handled is discussed in Methods. Then multiple linear regression will be calculated to find the coefficients of the independent variables, those that we are researching, and the interactions found in Results. Next these results will be analyzed to answer the research question, and the results will be discussed in the context of previous research and on improvements that can be made. The latter is found in Discussion

Theoretical framework

Human capital, what is that and what are its components - Human capital refers to the idea that production is increased when people invest in themselves. People can do this through furthering their education, training, or practicing a trade and staying healthy (Goldin 2016). By increasing the production, earnings of an individual increase respectively (Mincer, Polachek 1976).

It is presumed that parental education is positively correlated with the education of a child (Nelson, 2009) because people with higher education tend to influence their children to pursue higher education in their lives. As higher education leads to an increase in human capital, it is expected that this will cause higher earnings as well. Therefore, there might be an indirect correlation between education of parents and earnings of a child, with education of a child as a proxy variable. This is not a bidirectional interaction as it can safely be assumed that a child will not be paying for their own parent's education and so their earnings will have no effect on the education of their parents.

It is shown (Cave, Wright, von Stumm 2022) that education of parents is correlated with IQ of their children. From the data from 2000-2002, the coefficient of correlation of mother's education on children's IQ is 0.23 and the coefficient of correlation of father's education on IQ of a child was 0.20 with p value of 0.018. Since $p<0.05$, it can be assumed that correlation is of statistical significance. Zax and Rees have shown that a decrease of 15 IQ points results in an 11% decrease in earnings. Therefore, it is safe to assume that there is an indirect causal relationship between the education of parents and earnings of their child with the proxy being IQ. Unfortunately, IQ score cannot be found in the NLSY97 database and therefore ASVAB test score is used, since ASVAB and IQ try to measure the same thing, general cognitive ability.

Relationship between education of a mother and health - In a study (Güneş 2015) it was shown that maternal education shows significant correlation with the health of their child. This was demonstrated by there being a negative correlation between education of mother and their child having an extremely low birthweight (coefficient of -0.034 and $p<0.01$). Furthermore, this notion is pushed even further when looking at the correlation between maternal education and the height of their child which had a strong positive correlation (coefficient of 1.088 and $p<0.05$). As health is a component of human capital, it is fair to assume that health has a positive impact on earnings (even more so as a healthier person has the capability to be more productive). Therefore, there is an indirect causal relationship between the education of a mother and the earnings of their child, with the proxy being the child's health.

Control variables – Gender is consistently proven to have a substantial influence on the earnings of an individual (Weichselbaumer & Winter-Ebmer, 2003). Race also plays a role in determining one's earnings (McCall, 2001). What these variables have in common is that parental education plays no role in determining them, a parent's education cannot determine the

race or gender of their child. Therefore, these variables should be held as controls to help determine more accurately the effect that parental education has on one's earnings.

Methods

Dataset used in this research report is the national longitudinal survey of Youth 1997. Dataset consists of answers given by 8984 respondents. 39 variables were measured through two personal and telephone interviews with people from the US born between 1980 and 1984. The first interview took place when the respondents' ages ranged from 12 to 18 and the second when they were aged 34 to 40. In this dataset variables can have both negative and non-negative values. Negative numbers suggest reasons for not having an answer to the questions and non-negative values mean that an answer to the question is given. Therefore, we decided to extract only observations with non-negative values for variables we use in our research. After this operation we obtained a new dataset called (newdata) which consists of 3748 observations. Criteria used: (hdegparent>0 & hdegree>-1 & ASVAB>0 & wagealt>0 & whours>0 & tenure2017>=0 & height2002>0 & weight2017>0 & health2017>0)

All variables were taken from the newdata dataset, consisting of variables/interaction variables: HDegparent, HDegree, ASVAB and health2017. Also, the control variables: Tenure2017, race and gender were used. And lastly, the dependent variable of the model, wage. Wage: measured as continuous data with data ranging from ($0)-($9,999,999), measured to 2 decimal places. The values represent the respondent's hourly rate of pay multiplied by their hours worked per week. Finally, the natural log is taken of the wage which then turns into Ln(wage).

These concepts are measured by the following methods. HDegparent: was initially measured on a scale from 1 to 8, where the numbers are a scale of the highest degree of education received by one of their parents, ranging from no education (1) to professional degree (8). But for this data to be properly usable in a multiple linear regression the variable needed to be split into 8 different dummy variables consisting of: NoneEducationPar, GEDEducationPar, HighschoolEducationPar, AssociateEducationPar, BachelorEducationPar, MasterEducationPar, and PHdProfessionalEducationPar. These dummies can take values of either 1 or 0, 1 meaning that level of education was the highest degree attained by the respondent's parent and 0 meaning that level of education was not the highest degree attained by the respondent's parent. HDegree: Initially in the nlsy97 dataset, the highest degree achieved data spanned from 0 to 7, where (0) is no education and (7) is professional degree, for the same reason as for (HDegparent), this variable needed to be split into 8 different dummy variables consisting of: NoneEducation, GEDEducation, HighschoolEducation, AssociateEducation, BachelorEducation, MasterEducation, and PHdProfessionalEducation. These dummies can take values of either 1 or 0, 1 meaning that level of education was the highest degree attained by the respondent and 0 meaning that level of education was not the highest degree attained by the respondent. ASVAB: measured as continuous data with data ranging from 0 to 100,000 where 1 unit represents 0.001% ASVAB test result, thus, ranging from 0-100%. Health: measured as 5 different dummy variable on the respondent’s opinion consiting of: VPoorHealth, PoorHealth, GoodHealth, VGoodHealth, ExcellentHealth. These dummy variables take the value of either 1 or 0, 1 to say the respondent is of that health status and 0 meaning they are not. Tenure2017: tenure measured in weeks spent working at the respondent’s current employer ranging from 0 to 2,000 weeks. Race: as race is not quantifiable it is split into 4 dummy variables, consisting of: black, Hispanic, mixed, and white. These dummy variables take the value of either 1 or 0, 1 to say the

respondent is of that race and 0 meaning they are not. Gender: similar to race, gender is not quantifiable and therefore is split into 2 dummy variables: male and female. The dummy variables can take values of 0 and 1, 1 to signify the respondent is that gender and 0 to signify the respondent is not of that gender.

In this paper multiple linear regression will be used with LogWage as the dependent variable and the highest received degree of the parents as the independent variable. With the following interaction variables: ASVAB, Highest Degree received, and health. Also, the control variables: tenure, race, and gender. One of each dummy variable is excluded from the model as if all the other corresponding dummies have a value of 0 it means the same as the one dummy having a value of 1. These dummies chosen to be excluded are: NoneEducationPar, NoneEducation, Female, White, and ExcellentHealth. These were chosen due to their big sample sizes.

$$\hat{Y} = \beta_0 + \beta_1(\text{GEDEducationPar}) \beta_2(\text{HighschoolEducationPar}) + \beta_3(\text{AssociateEducationPar}) + \beta_4(\text{BachelorEducationPar}) + \beta_5(\text{MasterEducationPar}) + \beta_6(\text{PHD\&ProfessionalEducationPar}) + \beta_7(\text{GEDEducation}) + \beta_8(\text{HighschoolEducation}) + \beta_9(\text{AssociateEducation}) + \beta_{10}(\text{BachelorEducation}) + \beta_{11}(\text{MasterEducation}) + \beta_{12}(\text{PHD\&ProfessionalEducation}) + \beta_{13}(\text{ASVAB})_i \beta_{14}(\text{Male}) + \beta_{15}(\text{Black}) + \beta_{16}(\text{Hispanic}) + \beta_{17}(\text{Mixed}) + \beta_{18}(\text{Tenure2017})_i + \beta_{19}(\text{VGoodHealth}) + \beta_{20}(\text{GoodHealth}) + \beta_{21}(\text{PoorHealth}) + \beta_{22}(\text{VPoorHealth}) + \varepsilon$$

Results

Table 1 shows measures of central tendency and variation using the mean and standard deviation of all the regressors used in the regression analysis.

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.
LogWage	3,748	11.19	0.89
hdegparent3	3,748	3.70	1.52
ASVAB1	3,748	49.48	29.10
hdegree1	3,748	3.838	1.49
tenure2017	3,748	255.73	230.42
health2017	3,748	2.27	0.92
male	3,748	0.48	0.50
white	3,748	0.56	0.50

Table 1 shows the mean, standard deviation, and number of values for all 3748 respondents answers to the interviews for questions concerning the regression. Majority of respondents of the survey were female and white. Tenure has a relatively large standard deviation which can be explained by the large spread. Moreover, the mean of education of respondents was higher than the mean of the highest degree received by their parents, which means that respondents tend to have higher degrees than their parents, this is probably due to the fact that pursuing further education becomes more common as time goes on. Furthermore, people in this sample tend to have health closest to Poor.

Multiple linear regression was used to test if the highest degree of the parents significantly predicted wage. The regression model used was:

$$\hat{Y} = \beta_0 + \beta_1(\text{GEDEducationPar})\ \beta_2(\text{HighschoolEducationPar}) + \beta_3(\text{AssociateEducationPar}) + \beta_4(\text{BachelorEducationPar}) + \beta_5(\text{MasterEducationPar}) + \beta_6(\text{PHD\&ProfessionalEducationPar}) + \beta_7(\text{GEDEducation}) + \beta_8(\text{HighschoolEducation}) + \beta_9(\text{AssociateEducation}) + \beta_{10}(\text{BachelorEducation}) + \beta_{11}(\text{MasterEducation}) + \beta_{12}(\text{PHD\&ProfessionalEducation}) + \beta_{13}(\text{ASVAB})_i\ \beta_{14}(\text{Male}) + \beta_{15}(\text{Black}) + \beta_{16}(\text{Hispanic}) + \beta_{17}(\text{Mixed}) + \beta_{18}(\text{Tenure2017})_i + \beta_{19}(\text{VGoodHealth}) + \beta_{20}(\text{GoodHealth}) + \beta_{21}(\text{PoorHealth}) + \beta_{22}(\text{VPoorHealth}) + \varepsilon$$

(Refer to the appendix for Table 2)

Table 2. shows that the regression was indeed statistically important with an R-squared and adjusted R-squared of 0.197 and 0.202 respectively, also with an F-statistic of 42.86 on 22 variables and 3725 DF.

Furthermore, because of LogWage being a logarithmic function, we can interpret the values of coefficient as the percentage of change. Thus, for example, people whose parents' highest degree of education was GED earn (-0.142*100%=-14.2%) less than people in base group, being people whose parents' highest degree of education was none.
Without any control or interaction variables, R-squared for the regression with parental education as independent variable was 0.04 and has increased to 0.137 after adding interaction variable of one's education. Therefore, it means that one's education is more strictly correlated to the logWage than parental education.

Moreover, absolute values of coefficients of correlation of parental education have decreased after introducing interaction and control variables. For example, in model (0) people whose parents' highest degree was PhD or Professional earn (0.708*100% = 70.8%) more than people in base group, being people whose parents' highest degree was none. After introduction of interaction and control variables in model (6) people whose parents' highest degree was PhD or Professional earn (-0.003*100% = -0.3 %) less than people in base group, being people whose parents' degree was none. Similar trends can be seen for all dummy variables of parental education.

It was also found that ASVAB was able to predict the wage less well, but keep in mind that ASVAB has scores ranging from 1-100 thus the coefficient of 0.003 is still significant whereas ASVAB also has a P-value of under 0.01. Tenure has surprisingly little impact on wage but can be explained due to its large range, much like ASVAB. Moreover, surprisingly it was discovered that tenure has no significant influence on the LogWage(coefficient being 0.000).

Discussion

In conclusion, the highest significant degree of a parent does not have a noteworthy influence on the earnings of a child. After carrying out multiple regressions and adding more variables to the equation, significance of highest significant degree has been gradually decreasing. Education seems to be one of the best indicators of earnings. Moreover, education seems to some extent hereditary, since people with higher education tend to have parents with higher education.

These results greatly differ to a study from 2011 (Hudson, Sessions 2011) where there was thought to be exactly 0.00 correlation between parental education and log hourly wage (coefficient: 0.00, T-stat: -0.48). This is probably due to the difference in time period in which the results were found. (Hudson, Sessions 2011) took data from the US National Longitudinal Survey of Young Men from 1966-1981 and the data from this paper was from 2002-2017. It has become much more common over the last few decades for people to pursue a higher education and so there will be more data to provide evidence for correlation than there would have been between 1966 and 1981, especially when it concerns one's parents. Moreover, difference can be caused by the difference in multiple linear regression models used.

Various implications can be derived from this research. For example, if the sole purpose of the government was to increase the wages of further generations, then it may not be the most optimal decision to invest more money into education. This does not mean that the government should decrease funding of education as it still has a large impact on the wages of present generations. Furthermore, if one's objective was to maximize their earnings, the best way to fulfill this objective is to get a better education, as can be seen in model (6) in Table 2. People with higher education tend to have higher earnings. This implication is in line with the human capital theory.

A limitation of this research is that data for parental education is qualitative and therefore is not as accurate as if it were to be continuous. Furthermore, the model is not perfect as it assumes a bachelor's from one university is exactly equal to another, for example, a bachelor's from Arizona State University is equal to one in the same course at Harvard. More than this, a degree in one course is not perfectly equal to a degree in another as they can be vastly different areas like science and art, and yet, the model takes them as equals.

A suggestion for further research could be pedagogical research which could try to find theoretical evidence for or against the idea that higher educated parents are better at raising their children. Moreover, education in the USA differs greatly from other countries in the world since its' quality depends on the individual's wealth. Therefore, it is not advisable to map results of this research on the population of the entire earth. Research can be extended in other countries in which education is free or significantly cheaper than in the USA and thus, become more universal.

Bibliography

Becker, Gary S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to education* (Third ed.). Chicago: The University of Chicago Press.

Card, D. (1999). *Handbook of Labor Economics Volume 3.* Amsterdam: North-Holland Publishing Company. Pages 1801-1863.

Cave, Sophie N., Wright, Megan, & Stumm, Sophie von. (2021). *Change and stability in the association of association of parents' education with children's intelligence.* Sciencedirect. doi:https://doi.org/10.1016/j.intell.2021.101597

Goldin, Claudia (2016). *Human Capital.* Harvard Library. Retrieved from https://dash.harvard.edu/handle/1/34309590

Güneş, Pinar M. (2014). *The role of maternal education in child health: Evidence from a compulsory schooling law.* Sciencedirect. doi:https://doi.org/10.1016/j.econedurev.2015.02.008

McCall, Leslie (2001). Sources of Racial Wage Inequality in Metropolitan Labor Markets: Racial, Ethnic, and Gender Differences. *American Sociological Association*. Retrieved from https://www.jstor.org/stable/3088921

Nelson, Julie K. (2009). *Impact of parent education on student succes.* Retrieved from https://files.eric.ed.gov/fulltext/ED507263.pdf

John Hudson, John G. Sessions (2009) .Parental education. Sciencedirect. doi:https://doi.org/10.1016/j.econlet.2011.05.053

Schultz, Theodore W. (1974). *Marriage, Family, Human Capital, and Fertility.* Journal of Political Economy 82(2), Part II, April 1974. Retrieved from http://www.nber.org/chapters/c3685

Weichselbaumer, Doris, & Winter-Ebmer, Rudolf. (2005). A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys*. Retrieved from https://www.researchgate.net/publication/4911992_A_Meta-Analysis_of_the_International_Gender_Wage_Gap

World Health Organization. (2005). *The SuRF Report 2.* Geneva. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/43190/9241593024_eng.pdf

Appendix

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.
LogWage	3,748	11.19	0.89
hdegparent3	3,748	3.70	1.52
ASVAB1	3,748	49.48	29.10
hdegree1	3,748	3.838	1.49
health2017	3,748	2.27	0.92
tenure2017	3,748	255.73	230.42
male	3,748	0.48	0.50
white	3,748	0.56	0.50

Table 2: linear regression of LogWage based on the NLSY97

	Dependent variable:						
	LogWage						
Model	(0)	(1)	(2)	(3)	(4)	(5)	(6)
GEDEducationPar	−0.139 (0.090)	−0.154 (0.085)	−0.187∗∗ (0.085)	−0.191∗ (0.085)	−0.176∗∗ (0.084)	−0.150∗ (0.082)	−0.142* (0.083)
HighschoolEducationPar	0.108∗∗ (0.051)	-0.003 (0.049)	−0.036 (0.049)	−0.045 (0.050)	−0.035 (0.049)	−0.045 (0.048)	−0.033 (0.050)
AssociateEducationPar	0.186∗∗∗ (0.061)	0.029 (0.060)	−0.083 (0.060)	−0.091 (0.060)	−0.075 (0.060)	−0.084 (0.058)	−0.072 (0.060)
BachelorEducationPar	0.345∗∗∗ (0.057)	0.038 (0.060)	−0.038 (0.058)	−0.047 (0.058)	−0.030 (0.057)	−0.065 (0.056)	−0.053 (0.059)
MasterEducationPar	0.481∗∗∗ (0.066)	0.082 (0.066)	−0.007 (0.067)	−0.019 (0.067)	0.002 (0.067)	−0.048 (0.065)	−0.037 (0.068)
PHDProfessionalEducationPar	0.708∗∗∗ (0.091)	0.160 (0.091)	0.054 (0.092)	0.019 (0.092)	0.054 (0.091)	−0.013 (0.089)	−0.003 (0.091)

GEDEducation	0.180	0.147∗	0.139	0.133∗	0.140∗	0.143*
	(0.078)	(0.078)	(0.078)	(0.077)	(0.075)	(0.076)
HighschoolEducation	0.277*	0.214∗∗∗	0.205∗∗	0.164∗∗	0.198∗∗∗	0.199***
	(0.070)	(0.070)	(0.070)	(0.070)	(0.068)	(0.068)
AssociateEducation	0.245***	0.250∗∗∗	0.239∗∗	0.219∗∗∗	0.299∗∗∗	0.299***
	(0.080)	(0.081)	(0.081)	(0.080)	(0.079)	(0.079)
BachelorEducation	0.700***	0.570∗∗∗	0.549∗∗∗	0.504∗∗∗	0.591∗∗∗	0.593***
	(0.074)	(0.076)	(0.077)	(0.076)	(0.075)	(0.075)
MasterEducation	0.893***	0.744∗∗∗	0.727∗∗∗	0.677∗∗∗	0.810∗∗∗	0.812***
	(0.081)	(0.084)	(0.084)	(0.083)	(0.082)	(0.082)
PHDProfessionalEducation	1.460***	1.276∗∗∗	1.250∗∗∗	1.243∗∗∗	1.346∗∗∗	1.351***
	(0.106)	(0.110)	(0.110)	(0.109)	(0.107)	(0.108)
ASVAB1		0.004***	0.003∗∗∗	0.003∗∗∗	0.003∗∗∗	0.003***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VGoodHealth			-0.040	−0.035	−0.023	-0.024
			(0.036)	(0.035)	(0.035)	(0.035)
GoodHealth			-0.057	−0.047	−0.022	-0.023
			(0.036)	(0.038)	(0.037)	(0.037)
PoorHealth			-0.151	−0.124∗∗	-0.104*	-0.103*
			(0.056)	(0.056)	(0.054)	(0.054)
VPoorHealth			-0.245	-0.180	-0.0725	-0.074
			(0.150)	(0.149)	(0.146)	(0.146)
tenure2017				0.001∗∗∗	0.000***	0.000***
				(0.0001)	(0.000)	(0.000)

male						0.353*** (0.026)	0.351*** (0.027)
black							−0.026 (0.035)
hispanic							0.012 (0.038)
mixed							−0.108 (0.132)
Constant	10.991*** (0.046)	10.716*** (0.071)	10.680*** (0.072)	10.487***(0.071)	10.651*** (0.077)	10.440*** (0.077)	10.451*** (0.081)
Observations	3,748	3,748	3,748	3,748	3748	3748	3,748
R2	0.04	0.137	0.145	0.147	0.1637	0.2016	0.202
Adjusted R2	0.04	0.134	0.142	0.145	0.1597	0.1976	0.197
Residual Std. Error	0.869 (df = 3741)	0.824 (df = 3735)	0.820 (df = 3734)	0.820 (df = 3733)	0.8119 (df = 3729)	0.7933 (df = 3728)	0.794 (df = 3725)
F Statistic	25.7 (df=6; 3741)	40.29*** (df = 12; 3735)	48.785*** (df = 13; 3734)	37.91*** (df = 17; 3733)	40.55*** (df = 18; 3729)	49.56*** (df = 19; 3728)	42.86*** (df = 22; 3725)

Note: *p<0.1; **p<0.05; ***p<0.01; df = degrees of freedom; values in brackets = standard deviation.