The effect of different educational levels and institutions on an individual's salary.

535795 Joakim Isoaho

535889 Tuomo Juntunen

515553 Totti Tyynilä

1. Introduction

In this paper, we will study the effect of different educational levels and institutions on an individual's current and future salaries. One of the great decisions a person makes in life, is choosing the right education for maximizing their future utility. Investment into education and future salaries are crucial factors in maximizing utility. Utility and well-being are often measured in economics by money, which is why we want to study the progression of salaries as the education variable changes. We cannot determine individual's subjective utility function, but a rational individual must consider the effect of a higher degree on the future salaries, and adjust that to one's preferences. In this paper, we will examine the effect of different educational levels on the salary of an individual.

The research question isn't relevant only from the deciding individual's point of view, but also from the society's perspective. Especially in Finland, where education is provided and subsidized by the government, it is essential to understand the outcomes of different degrees of education. From the society's perspective, it is important to study the education's effect on salaries, as the future education is again funded by the taxes gained from people's future salaries.

The research question is relevant whether it is studied in the U.S. or Finland, although the perspective changes financially from the United States of America's individual point of view to Finland's more collective point of view. Finland's recent budget cuts on education make the research question even more relevant. Is it smart to reduce education spendings if gains from increased higher education bring in more tax funds? Not to mention other positive externalities. The research question is vital as it applies to virtually every individual and offers necessary perspective on whether investing in education is the right decision for maximizing utility.

This research topic has been studied extensively in the past and practically every one of them supports our intuitive hypotheses: higher education generates higher earnings. For example, Carnevale, Cheah and Rose (2009) state that "no matter how you cut it," more education pays, but earnings vary greatly depending on the degree type, age, gender, race/ethnicity, and occupation of an individual. The 33 percent of Bachelor's degree holders that continue on to graduate have even more prosperous futures ahead. Moreover, the difference in earnings between those who go to college and those who don't is growing; meaning that postsecondary education is more important than ever (Carnevale, Cheah & Rose, 2009). Also, Oreopoulos and Petronijevic (2013) stress the growing importance of post-secondary and post-baccalaureate education. They also state that earnings benefits also appear to be associated more with some college majors than with others (Oreopoulos & Petronijevic, 2013).

2. Data & Theory

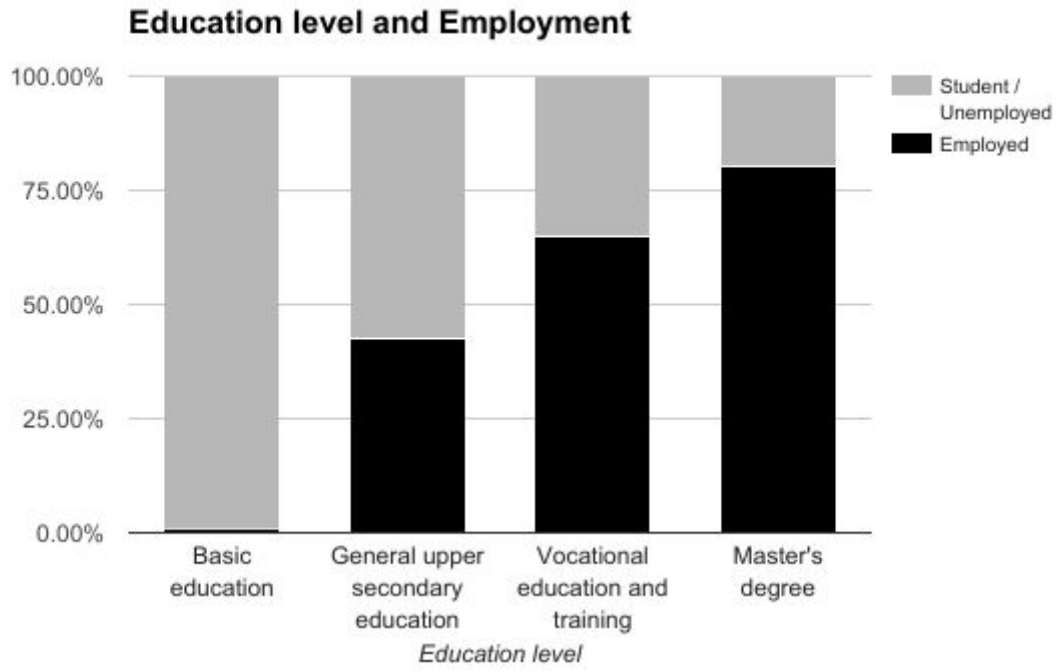
In order to analyze the effects of different education levels on future salaries, we needed to find the correct sources of data. Through some searching on multiple different information services, multitudes of different data sources emerged. Most of these were however in the form of already processed charts or narrow tables of data, which are not fit for our analysis. On top of that most results were based in USA, which was not in the scope of our research question.

Through some more in-depth research we found two critically important data sources. The first would be used as auxiliary evidence for our hypothesis, and the second as the source of the data for the actual econometric model.

1. **Vipunen.fi**, which has statistics based on data and registers collected by the Statistics Finland, the Ministry of Culture and Education and the Finnish National Board of Education. As the data is sources by a governmental entity, it can be generally regarded as trustworthy. However, one could point out a potential bias *for* education because of the connection to the Ministry of Education.

2. **Statistics Finland** (Tilastokeskus), the national statistical institution in Finland, established in 1865 to serve as an information service and to provide statistics and expertise in the statistical sciences. Regarded as a trustworthy source and has been cited multitudes of times in finnish research papers.

Through Vipunen.fi we could utilize pre-aggregated datapoints about the placement of students after completion of a certain level of studies. The different levels of studies it uses (In order by least years of education) are: Basic education, General upper secondary education, Vocational education and training, Bachelor's degree, Master's degree. The datapoints are surveyed one year after completion of the study level. This allows for more accurate results on actual placement to be gathered.



As this is only auxiliary data, we will inspect it only lightly. Basic education having so little employed is because of the fact that nearly every single person in Finland continues as a student after completing it. However, a certain trend can be realized: Higher education level (=more time spent on education) seems to be leading to higher employment rates. This result however cannot be taken at face value as the master's degree is only 7.24% of the sample's n, while Vocational Training is 30.25%.

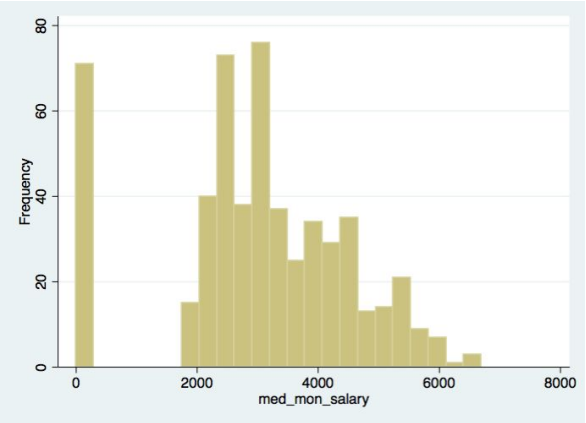
Private sector Monthly Salaries seperated by Education Level, 2015								
Education Level	Years of Education	Sample Size			Salary, €/month			
		Sum	Male	Female	Average	1st Decile	Median	9th Decile
Basic Education	9	56 601	32 857	23 744	3 035	1 933	2 631	4 603
Keskiaste	11.5	238 097	120 492	117 605	2 928	1 983	2 622	4 215
Alin korkea-aste	14	110 069	44 757	65 312	3 632	2 350	3 276	5 298
Bachelor's Degree	17.5	156 242	86 982	69 260	3 638	2 343	3 333	5 294
Master's Degree	23	124 796	66 425	58 370	4 712	2 809	4 281	7 048
Doctoral or Licensiate	28	15 937	9 057	6 880	5 184	3 482	4 799	7 235
Sample		701 742	360 571	341 171	3 574	2 127	3 170	5 486

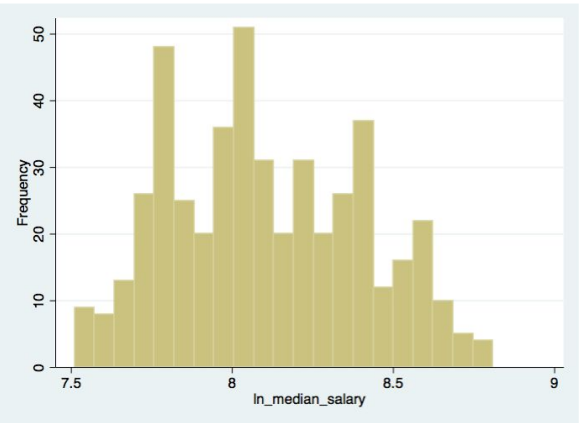
A summary table aggregated by Statistics Finland[1]

The actual data used was constructed in the following way: Through the Statistics Finland database we gain access to different types of relevant data. Unfortunately it is still quite unsatisfactory, as it does not include individual observations, but only aggregated monthly salaries of different groups. Thus we have to take some steps that otherwise would not have needed. We need to be aware of issues, such as the Ecological Fallacy, and realize that we do not for example have information about the skewness of the dataset. This is a huge problem.

What we can do however, is use what we have and calculate some of the missing characteristics through information we do have. We have the coefficient of variation, and data of the 1st and 9th decentile. These can be used to gather information about the skewness, and also to calculate the standard deviation. We also can utilize general information we know about the nature and distribution of salaries.

The first thing to note is the use of median salary instead of mean salary. This is largely because of the outlier effect on the mean: there are large amount of observations on either end of the data which inherently skew the mean past the point of usefulness. In these instances, the median is much more useful since it gives a more accurate representation of the middle of the pack than the mean. But even with the median salary, we still need to process it further.





As we can see, the median salary distribution is truncated at 0 and is highly right-skewed. If you try to use just the absolute salary to estimate the effects of various things on salaries, you'll probably end up with people making estimated negative salaries, which not at all what we are looking for. That is why

[1] *(Education levels are not official translations)*

we'll be using the natural logarithm of the median salary. It makes more sense to measure "percent" changes in salary (which is what ln(salary) is), rather than absolute changes in salary. Also, the histogram of this starts looking more like normal distribution which will help us in the future.

Our main explanatory variable is the education level, which we changed from a categorical string variable to an ordinal integer variable. The values start from 1 (Basic education) and end in 6 (Doctorate).

The database also includes gender information, which we encoded to be a dummy variable "female", gaining the value 0 if male, and 1 if female. Next up are age-group categories, going from under 20 years old (-20v) to 65 years and older (65+). For that we've created an ordinal variable for each group ranging from 1-9.

Age Groups:	-20v	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+
Identifier	1	2	3	4	5	6	7	8	9	10	11

Our main Hypothesis is that high education levels lead to higher monthly earnings and thus higher average lifetime earnings. Even without consulting earlier literature there are strong theoretical reasons to believe it is right. We base our thoughts on the hedonic assumption that companies pay individuals based on their output. This output, as is common in economic literature, we assume is largely based on the aspect of productivity and efficiency. We then assume that the key driving factor behind productivity and efficiency is knowledge. Education, as Argote, L. & Ingram, P. (2000) put it, is the transferring of knowledge to others. Thus it brings us to the logical conclusion that Education should have a significant effect on salary.

3. Analysis

A simple model

As we stated in the introduction of the paper, we set out to find how differences in education might affect the salary of an individual. Framing this in an econometric setting, what we want to get out is a mathematical function that might look something like this:

$$Salary = f(Education)$$

Salary is the dependant variable, or in other words the outcome to be explained. Education is the explanatory variable, or in other words a factor that explains the outcome. Lastly, the function f translates education into monthly salaries.

```
. reg ln_median_salary education_level
```

Source	SS	df	MS
Model	14.7682999	1	14.7682999
Residual	25.6063409	460	.055665958
Total	40.3746408	461	.087580566

Number of obs	462
F(1, 460)	265.30
Prob > F	0.0000
R-squared	0.3658
Adj R-squared	0.3644
Root MSE	.23594

ln_median_sal~y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
education_level	.1075663	.006604	16.29	0.000	.0945886	.120544
_cons	7.713153	.0260797	295.75	0.000	7.661903	7.764403

Doing a regression only with the natural logarithm and the education level. We observe a positive coefficient. This is with 95% confidence interval and thus quite believable.

More accurately our final model will be based on something akin to this:

$$ln(Salary) = f(Education, Gender, Age)$$

$$ln(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Potential other models

After multiple days of researching about labor economics we constantly read mentions about the landmark work of Jacob Mincer, author of Schooling, Experience and Earnings (Mincer, 1974). On basis of both theoretical and empirical arguments, Mincer modeled the natural log of salary as a function of years of education and years of potential experience. The most widely used version of "Mincer's equation looks like this:

$$ln\ \omega = ln\ \omega_0 + pS + \beta_1 x + \beta_2 x^2$$

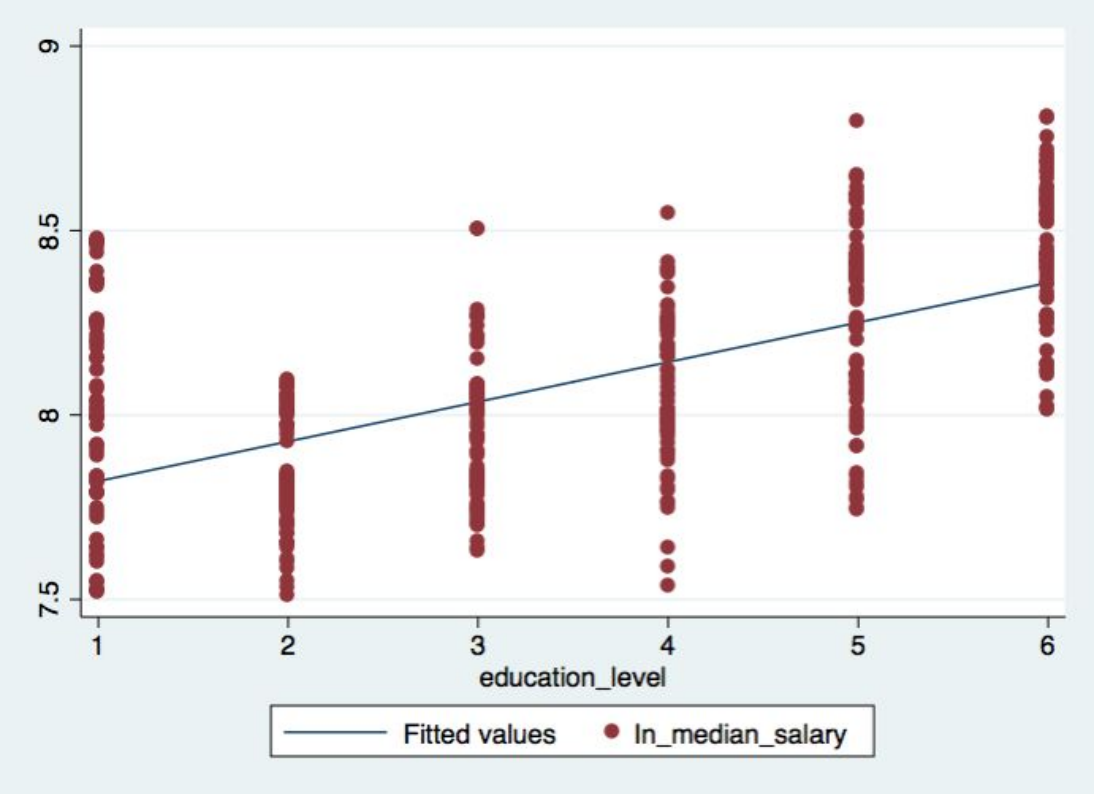
Here ω is earnings (With ω_0 being the salary of a person owning no education or experience), S is years of education, and x is years of potential experience.

After multiple revisions the results of this model is as follows: p, the rate of return on education is estimated to be around 5-8% per year, but can range from as low as 2% to as high as 20%. One very important thing to note here is how Mincer has handled the introduction of experience to the equation. In our model we simply had a categorical variable indicating general age group. This is highly problematic as it naturally correlates with education level.

The other problem we had was how years of experience ($\approx$age) has drop-offs at both ends: very young and very old people earn less than people in the middle. This is why in Mincer's model the experience term is quadratic. It usually takes the values of $\beta_1 \approx 0.03$ and

$\beta_2 \approx -0.004$. This allows the model to have a concave effect. The return on experience begins positive, and then falls later on. The whole model is actually pretty genius, and can include for potential dummy variables as well. You can control for things such as gender, skin type or ethnicity.

4. Conclusions



Variable	a1	a2	a3
education_~l	.10716507	.10720425	.10718966
female		.0043889	.00431471
age_category			-.00075093
_cons	7.715054	7.7127383	7.7172614

$$ln(Y) = 7.7172614 + 0.10718966x_1 + 0.00431471x_2 - 0.00076093x_3$$

The linear regression shows that our main hypothesis turned out to be accurate: higher levels of education do tend to correlate with higher levels of monthly earnings. What we can see from the results is that the base salary for someone with only the basic education level is ln 7.7172614. Then, by start off with just the education level variable and see it having the positive value of 0.10716507. In general terms this means that a one level change in education level has a 10.7% positive effect on median salary. The effect continues to be on the same levels throughout the adding of additional variables.

We can see that gender does in fact play some role in future earning levels: women hold a slightly lower salary level than men. The gender pay gap (GPG) in Finland for <25 year olds is at 3,4% which is drastically lower than that of the 25-34 age group, in which the GPG is at 11% (Eurostat, 2017) A possible reason for this can be that men and women tend to choose different career options already when choosing their education (Haapasalmi, 2013). If we assume that certain professions have different levels of remuneration although they might require a similar level of education, the addition of a gender component would skew our research because men and women do different kinds of jobs.

Economic and statistical significance

The economic significance of the study is to further provide evidence that education can be seen as an investment into one's future earnings. The higher level of education an individual has, the higher the one's future annual salary can be projected to be.

Due to the use of the median salary instead of an average salary, the effect of outliers is diminished, giving the data robustness. We also decided to use the natural logarithm of the median salary in order to make the results comparable. When we can see from the STATA report that the R-squared value for our data is 0,3124, which is a relatively low score. A point of interest can be seen in figure 4.2. There is greater variance in the education level 1 (basic education) data than any other data set. A possible explanation to this is that our model can only take into consideration the formal forms of education. It cannot account for, for example, craftmanship that doesn't require formal education, but can only be accrued only through years of tradecraft and is accordingly compensated. In addition to this explanation, another possibility can be that some jobs do not require higher levels of education but can be unfavourable due to harsh working conditions, which can drive salary levels up.

What did we learn?

Our study showed that there is indeed a positive correlation between one's level of education. The intuition that one's salary tends to be higher when one completes higher levels of education might seem almost obvious to some. However, the knowledge our research provides is that different levels of education create different amounts of "added value." "Moving up" from secondary education to any form of higher education seems to create a rise in median salary higher than, for example, getting a Doctoral education instead of a Master's level education.

The initial idea of our study came from proposed budget cuts to the education sector suggested by the Finnish government. We did learn that investing in education, does cause higher levels of income and we can state that higher levels of income will correlate to higher amounts of tax. What our model shows is that especially cutting from secondary education would be detrimental to income levels, because cutting funding from the secondary level would most likely cause a drop in the levels of students seeking education past secondary level and lead to a lower income level and thus lower tax revenues. Thus, it is for the government to decide whether the savings acquired from cutting educational budgets will lead to greater benefits than the lower tax revenue caused by the cuts.

Potential sources of error/misinterpretation

A source of error in our study could possibly be the question of whether education in itself causes the observed rise in salary levels, or whether people who end up completing certain levels of education tend to have a certain level of income. Would the people seeking higher education be such that they would in any case be smart and productive enough to have higher salaries?

```
. regress ln_median_salary female education_level age_category if age_category < 7, vce(robust)

Linear regression                               Number of obs =     263
                                                F(  3,   259) =   33.43
                                                Prob > F      =  0.0000
                                                R-squared     =  0.3124
                                                Root MSE      =   .2369
```

ln_median_sal~y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	-.0110467	.0292402	-0.38	0.706	-.0686256	.0465322
education_level	.0950395	.0095419	9.96	0.000	.07625	.113829
age_category	.0063544	.0086142	0.74	0.461	-.0106084	.0233172
_cons	7.747618	.0527549	146.86	0.000	7.643735	7.851501

In our initial regression we observed female having a positive coefficient for the dependant variable, while age category had a negative coefficient. We tried coming up with explanation for this, and one of the main reasons we argued for was the suspected concavity of age on salary.

Doing the same regression again, but with age_category < 7 (under 49 years old) the results look far more intuitive. Females have a negative coefficient on salary, while education level and age both have positive coefficients. However, the P values are not very promising, and also the confidence interval shows a potential 0 result.

Omitted variables:

Socioeconomic status: Level of education tends to be inherited from parents to children. (Myrskylä, 2009) The ways parents view education has been shown to affect the way children view education and the parents' education tends to have an impact on the educational levels of their children. However, we considered this variable omittable, because we wanted to keep the focus on the effect one's own education had on their future income.

Ability: it can be assumed that one's ability will have a positive effect on both earnings and education it would be extremely difficult to take into account an individual's abilities because it is difficult to determine a measure of ability. One could suggest that IQ-levels could be used, but IQ does only test a limited scope of one's intelligence or abilities. This is why this variable has been omitted.

What could we study next?

A next step in our research could be to analyse how success in education correlates with future earnings. Our current research studies the relation between completing a certain level of education.

However, there is obviously some variation when it comes how one succeeds within a certain level of education. It could be interesting to study how grades correlate with future earnings. Intuition would say that those who succeed better in their studies are recruited to positions with better pay. This kind of a study would require more detailed data knowledge about the individual on scholastic success and future earning levels of the participants of the study that would require considerably more time resources than were available for our current study.

Possible Alternative scopes

Value for time used aspect: it could be interesting to study how much one's income levels increase compared to the time spent on education. Which level of education yield the greatest increase in salary per year spent in education?

Longer time frame: by taking a look at data from a longer time period we could see how the effect of the individual's level of education has changed over time. Would the addition of income brought by higher levels of education be significantly higher due to the fact that acquiring education was more rare historically?

Different cultures: it could be examined how education and salary are correlated in different cultures. In Finland education is free, but, for example, in the USA, especially private education is extremely expensive. It is not uncommon to take out significant sized loans for university education, because it is seen that the addition to one's salary due from getting a college degree is significant enough to pay for debts ranging from tens of thousands to hundreds of thousands USD.

References

Carnevale, A.P., Cheah, B., Rose, S.J. (2009). The College Payoff: Education, Occupation, Lifetime Earnings. The Georgetown University Center on Education and the Workforce. https://repository.library.georgetown.edu/bitstream/handle/10822/559300/collegepayoff-complete.pdf?sequence=1. Accessed: 24.4.2017

Clark, W. A., & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis*, 8(4), 428-438.

Eurostat (2017). Gender pay gap statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics . Accessed: 24.4.2017

Haapasalmi, J., Sajavaara, A. (2013). Miehet ja naiset työelämässä. https://ek.fi/wp-content/uploads/naisetmiehet_maaliskuu2013.pdf Accessed 25.4.2017

Myrskylä, P., (2009). Koultus periytyy edelleen. *Tilastokeskuksen Hyvinvointikatsaus 1/2009* http://www.stat.fi/artikkelit/2009/art_2009-03-16_002.html?s=0 Accessed 25.4.2017

Oreopoulos, P., & Petronijevic, U. (2013). Making College Worth It: A Review of the Returns to Higher Education. *The Future of Children, Vol. 23, No. 1, SPRING 2013*. http://files.eric.ed.gov/fulltext/EJ1015240.pdf. Accessed: 24.4.2017

Piantadosi, S., Byar, D. P., Green, S. B. (1988). The ecological fallacy. *American journal of epidemiology*, 127(5), 893-904.

Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes*, 82(1), 150-169.