Autumn, 2007
Empirical Methods 1
Institution for Economics
Stockholm University

Andreas Jantvik Gyllenlid

The direct effects of education on health

Content list

1. Introduction 1
   1.1 Purpose 1
   1.2 Question 1
   1.3 Overview of previous literature 1
2. Empirical strategy and data .... **Error! Bookmark not defined.**
3. Results 7
4. Discussions and conclusions 9

1. Introduction

Technological growth is by many thought of to be the sole long-run improver of economical conditions in the world. Education, and all incentives to educate oneself, should therefore be of outmost importance.

In this study, I examine the effects of education on health in Sweden. Being a student in this country, I find it important to define as many good reasons for higher education as possible. A more general study would have been to examine the effects of income on health, but I found this approach more interesting. Since education and income are obviously positively correlated, my mission in this study is to try to distinguish the education effect from all the other possible side-effects one gets from enduring higher education – such as higher future income.

It is my firm belief that a significant direct connection between a prolonged and healthier life and education may boost morale of students aware of this and may give entirely new incentives for higher studies. Although it is far from certain that this is the case, the mere possibility of such a prospect makes this worth to examine.

1.1 Purpose

The purpose of this paper is to study the effects of education on health in Sweden in a way that distinguishes this effect from other "side-effects" of studies.

1.2 Question

What is the effect of education on health in Sweden?

1.3 Overview of previous literature

In a report by ”Liv & Hälsa”[1], the authors find that there is a correlation between education levels and health. However, at the same time they write that this finding could be misleading. There are many other factors that can induce good health, such as working environments and income. They also find that people with no high school education are, on average, older than people with a high school diploma. The reason they presented for this difference was that higher education has become increasingly important over time. So while it was, perhaps, perfectly acceptable to start a working career without a high school education in the 1930:s, it most assuredly is less accepted today. This would lead to a higher average age of this

[1] ”*Livsvillkor, levnadsvanor och hälsa*”, Liv & Hälsa rapport nr 8, 2002

particular group of subjects, leading to omitted variable bias in favour of the effect of education on the health on lower educated people.

In another report presented in the American Sociological Review[2], the authors also argue that there are plenty side-effects of education, but they also mean that education does have a direct effect on health. Well educated people feel more in control of their lives are more sociable and practice less self-destructive behaviour. By enduring higher education, people tend to be endowed with more of the psychological resources required to better deal with life.

---

[2] ”The Links Between Education and Health”, C. Ross; Chia-ling Wu, *American Sociological Review*, Vol. 60, No. 5. (Oct., 1995), pp. 719-745

2. Empirical strategy and data

The models in this paper are based on the Swedish Level-of-Living Survey (LNU).[3] LNU is built on both retrospective collection of previous data and personal interviews. It is a random sample of people in the age of 16 to 74.

The dependent variable is "Y357" in the LNU dataset, which is a summed index of 44 health-related questions. Subjects are given points to each illness they perceive that they have. In this index they get zero points on a specific health-related question if they answer that they have no problem with that question, one point for a light problem and two points for a severe problem. The possible range total points for all 44 questions would thus be between 0 and 88. The independent variable of interest is total years of schooling ("Y131" in the LNU dataset). Variables to control for would be; some samples of destructive habits such as smoking ("Y340"), the frequency of hangovers ("Y345"). Of course, a given regressor of interest would also be the amount of income per month after tax ("Y473").

Since I want the effect of education on health, I would like to check for changes in the summed health index keeping personal things like smoking and income constant. This way I can see if education in itself has a direct effect on health.

I have chosen to convert the hangover variable and the smoke variable to binary variables. The hangover variable is constructed as such that all who report to ever get hangovers, be it every time they drink or only rarely, becomes 1. All who don't drink or never get hangovers becomes 0. Hangovers are generally a quite unpleasant experience, and getting hangovers would thus qualify as something self-destructive. I want to filter out the effects of various subject self-destructive behaviors on health and only get the education effect. Drinking with moderation is usually not considered to be an outright self-destructive activity, and so it is included on "the good side" of 0.

A somewhat different logic has been applied to smoking. Since hangovers from smoking normal cigarettes are possible, it is quite hard to achieve and most people never experience it. Thus the criteria for "self-destructive personality trait" are considered to be whether the person currently smokes at all or not. Smokers get 1, non-smokers get 0.

To conduct the test, I use regression analysis. Ordinary Least Squares (OLS) is used to compute the regression line.

---

[3] "The Swedish Level-of-Living Survey 1991", SOFI

A summary of the various variables and some descriptive statistics concerning them is given below.

Variable name	Description
Health	Summary index of health data
Education	Years of total schooling
Income	Monthly income after tax
Smoke	1 if smoker, 0 otherwise
Hangover	1 if hangovers occur, 0 otherwise

Variable name	Mean (Std.dev)	Min/Max
Health (n = 5231)	6,32 (5,58)	0/48
Education (n = 5300)	10,87 (3,41)	0/34
Income (n = 5179)	8590,96 (3305,72)	0/39999
Smoking (n = 5302)	N/A	N/A
Hangover (n = 5290)	N/A	N/A

3. Results

The first model I use is a bivariate regression model. It uses health as regressand and education as regressor.

$$Health = \beta_0 + \beta_1 \text{ x } Education + \varepsilon$$

$$Health = 9{,}41596 - 0{,}28703 \text{ x } Education, \ R^2 = 0{,}03035$$

$$(0{,}25628) \quad (0{,}02245)$$

The R^2 is quite low, so there is not much of the variation of health between various people that is actually explained by the model.

However, the t-statistic for β_1 is (-0 ,28703 – 0)/(0,02245) = -12,78530. We can see that the t-statistic in absolute value is large, so there is no doubt a correlation between the two. The hypothesis that $\beta_1 = 0$ and thus that it has no effect on health can easily be rejected at a significance level far below one percent.

The coefficient β_1 is negative. This means that education has a negative effect on the”Health points”-variable that Health is. Since a higher health-point score means more disease, we can conclude that education has a positive effect on health – or a negative effect on disease.

In the second model I add income as a regressor.

$$Health = \beta_0 + \beta_1 \text{ x } Education + \beta_2 \text{ x } Income + \varepsilon$$

$$Health = 9{,}280 - 0{,}2115 \text{ x } Education - (1{,}317 \text{ x } 10^{-4}) \text{ x } Income, \ R^2 = 0{,}04142$$

$$(0{,}2570) \quad (0{,}02434) \qquad\qquad (1{,}675 \text{ x } 10^{-5})$$

We can immediately see that the fit of the model has increased by a large share, about 25%, though it is still small in value. The difference is not simply due to the increase in the amount of variables, since the difference between the R^2 and the R^2-adjusted (0,04104) is negligible. Though we can say that predictions about health are more precise when Income is also a regressor, it is still quite imprecise.

The F-statistic for this test is 110,2. Thus we can easily reject the null that both Education and Income are zero. The t-statistic for β_2 is large, so we can also reject that β_2 would have zero effect on health.

β_2 also has a negative coefficient and thus income also affects disease negatively.

In the next and final model I include the regressors for self-destructive behaviour, smoking and hangover. I choose to only have three test-models included, because smoking and drinking represents the same characteristic entity to be examined.

$$Health = \beta_0 + \beta_1 \; x \; Education + \beta_2 \; x \; Income + \beta_3 \; x \; Smoking + \beta_4 \; x \; Hangover + \varepsilon$$

$$\underset{(0{,}2638)}{Health = 9{,}094} - \underset{(0{,}02450)}{0{,}2080} \; x \; Education - \underset{(1.692 \; x \; 10^{-5})}{(1{,}366 \; x \; 10^{-4})} \; x \; Income +$$

$$+ \underset{(0{,}1692)}{0{,}5675} \; x \; Smoking + \underset{(0{,}1645)}{0{,}03433} \; x \; Hangover, \; R^2 = 0{,}04398$$

The R^2 has changed only marginally, and so has the R^2-adjusted (0,04323). The F-statistic has sunk to 58.52, which is still a large value.

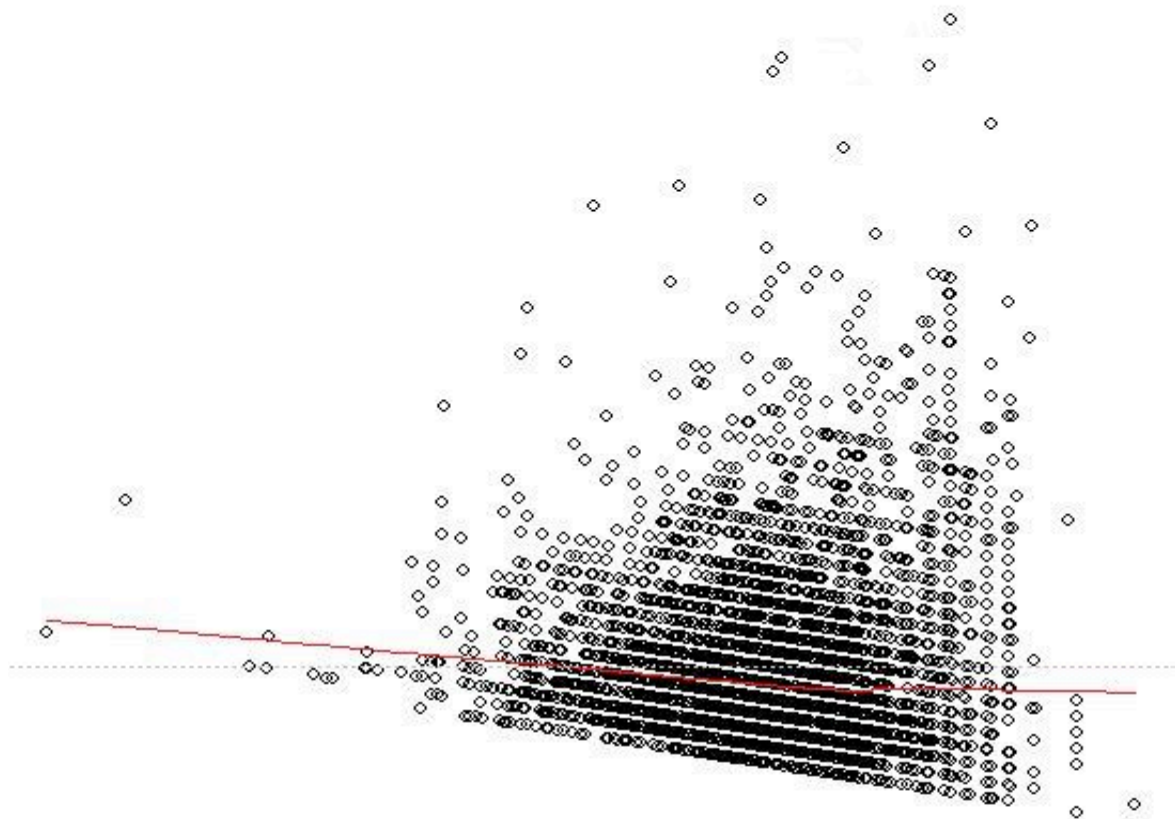
The t-statistic for β_3 is 3,35, and the same for β_4 is 0,21. The null hypothesis that $\beta_3 = 0$ can be rejected at the 1% significance level. However for β_4, we cannot reject the null even with a 10% significance level.

It seems like smoking and hangovers are positive to disease (and negative to health), however whether hangovers affect health negatively or not cannot be established using a 10% significance level.

4. Discussion and conclusions

I set out in this study to examine the effect of education on health. To do this I used the LNU dataset. I thought that to grab the pure effect of education on health, I should control for factors that are usually related to education. In this study, those factors were income and smoking- and drinking habits. I built models using the variables presented in the dataset, and after modifying some of the variables a little, I used regression analysis to conduct the tests.

The results were not at all as clear as I would have hoped, and a lot of the data did not meet my expectations. First of all, the R^2 and the R^2-adjusted were too small. This means that the outcomes of the health of the people in the sample are not very well described by my linear regression model. I looked at a scatterplot of the last model, and I think that it is likely that the fit of the model could have been better if a quadratic regression function had been used.



Scatterplot of model three.

However, none of the moments that I used as regressors were statistically insignificant, except for “hangover”. I chose treat hangover and smoking as “one variable” in the sense that I did not run separate tests for them. Considering that the variable hangover was statistically insignificant, I do not believe that this has affected my results in any non-negligible way.

After making this study, I can conclude that education has a positive effect on health. I can also conclude that this effect remains even when income and smoking is held constant. It is then sure that there is an education effect that is positive to health beyond the income effect. Actually, given the very small difference in β_1 between models one and two, we can conclude that the mayor part of the positive effect on health from education is not mainly due to the increase in income.
This is a strange result, because I definitely thought that the income effect would dominate over other effects that education may have on health. Like anyone else, I cannot exclude the possibility of error that would give a misleading result, however it seems quite certain that the major indirect effect education has on health does not, in fact, come from a higher income. Whether education has a direct effect on health cannot be known. Going through education has extremely many effects on a person and so it is possible – even likely – that I have simply missed the regressor that would have had the dominating effect on the regressand, health, that I was looking for.

In its goal to establish a definite direct connection between education and health, this study has failed. It now appears that such a question in itself is doomed to failure unless one has extreme resources to make sure that one finds the factors that dominates the indirect health effects and includes them in the regression model. After making this study, it seems as if my initial question was thought through a little diffusely.

Another point of critique is that, unfortunately, age was not a part of my dataset. I could not control for this variable, which means that there could be some omitted variable bias in favour of the effects of education on health on lower educated people (see section “1.3 Overview of previous literature”). Since the mean of the education variable is 10,87 with a standard deviation of 3,41, I think that this bias may have been relevant in my study. People without any high school diploma has studied at most 10 full years, and it would seem that plenty of subjects in my sample have studied only this much, and less.