SCHOOL OF
ECONOMICS AND
MANAGEMENT

Estimates of the Return to Schooling and Variations Between Countries and Ability Groups

By David Sandberg

Abstract: This study analyzes the return to schooling in 22 countries. More specifically, the effect on earnings of receiving a bachelor's and a master's degree is estimated by applying a propensity score matching approach on microdata. Unlike many other techniques, propensity score matching addresses selection bias to a higher extent making the estimates of the treatment effect on the treated more reliable. The findings suggest that receiving a degree results in around 15 percent higher earnings. To disentangle these estimates, each countries' return to schooling is estimated separately to examine the relation between the return and country-level variables where social expenditure and income equality are negatively related to the return. Furthermore, this study provides evidence that the return to schooling varies between individuals belonging to different parts of the ability distribution as low-ability individuals even have a negative return to a master's degree. The differences in returns across the ability distribution are even more evident when isolating countries with low social expenditure.

Keywords: return to schooling, selection bias, propensity score matching, cross-country analysis, ability

2021-05-26
Lund University
Department of Economics
Master's Degree (Two Years)
Supervisor: Andreas Bergh

Table of Contents

1. Introduction

Analyzing the return to schooling has interested economists since Mincer (1974) popularized and estimated it by using the model now referred to as the *Mincer earnings function*. This model uses education and work experience to explain earnings, but as richer microdata has become more accessible in recent decades, labor economists dig deeper into what determines the causal relationship between education and earnings by taking advantage of more detailed information. It is consensus that higher education does, on average, lead to higher earnings[1], but there is a need for a deeper understanding of the magnitude of the return, the cross-country variation, and the variation between different types of individuals.

In this study, I aim to find the return to schooling for almost 20,000 individuals across 22 countries. More specifically, I examine the effect on earnings of receiving a bachelor's and a master's degree by applying a propensity score model on microdata from the PIAAC survey from 2016. The findings suggest that the return of receiving a bachelor's or a master's degree results in around 15 percent higher earnings. Furthermore, I look separately into each country and examine how the return to schooling relates to four different country-level variables (education expenditure, social expenditure, income inequality, and income per capita). I find that a country's social expenditure is negatively, and income inequality is positively related to a higher return to schooling. This is reassured with intuition: if a country invests a lot of resources in welfare (high social expenditure), low-earning individuals tend to not fall too much behind the high-earning individuals, and a high return to schooling facilitates income inequality to grow.

I also make use of individual-specific ability, proxied by test score, to find how the return to schooling depends on where on the ability distribution one belongs. These results show a large variation between different ability groups where, for example, low-ability individuals are associated with a negative return to a master's degree. In addition, I examine the relation between the return to schooling and countries' social expenditure across the ability distribution. I find that countries with lower social expenditure tend to have a larger variation in the return to schooling across the ability distribution and a particularly high return for high-ability individuals.

1.1 Overcoming selection bias

Taking the average earnings of individuals with a degree and compare it with individuals without a degree says little about the effect of the degree since the two groups most likely differ in characteristics other than educational attainment. Intuitively, high-skilled individuals are more inclined to get a degree compared to low-skilled individuals resulting in that one cannot attribute the whole earnings difference between the two groups to their education level differences.

[1] In comprehensive reviews of the literature the cross-country average rate of return to a year of schooling is estimated to be approximately 10 percent a year and slightly higher for years when one earns a degree (Montenegro & Patrinos 2014; Psacharopoulos & Patrinos 2018).

Likewise, individuals with high-educated parents are also more inclined to attain higher education. This is called *selection bias*, which originates from sample selection or self-selection, and is often problematic when evaluating treatments.[2] A simple OLS regression has difficulties in addressing this bias as it assumes that treatment happens randomly and that all treated are affected similarly.

So, how can one isolate the effect of a certain education level on earnings and, thus, overcome selection bias? A common approach is to take advantage of exogenous determinants of education decisions by using an instrumental variable (IV) analysis. Examples of instruments used for educational attainment has, for example, been distance to school (Maluccio 1998; Kane & Rouse 1993; Card 1993), parental and spouses' education (Trostel et al. 2002), public tuition (Maluccio 1998), month of birth (Angrist & Keueger 1991; Leigh & Ryan 2008), and changes in compulsory schooling laws (Harmon & Walker 1995; Meghir & Palme 2003).[3] However, whether these IV approaches actually find the true return to schooling has been questioned by for example Heckman et al. (2006). The criticism is mainly directed against the strong assumption that all individuals respond in the same direction to the instrument and thus ruling out heterogeneous responses (Heckman et al. 2006; Heckman & Vytlacil 2005).

Also, the mentioned IV techniques primarily focus on the return to schooling years and, thus, overlook the potential signaling effect which, however, has proven difficult to distinguish from the human capital effect as the return to a degree is the combined effect of human capital accumulation and the signaling effect of being a graduate (Bedard 2001;).[4] One way to accurately estimate the signaling effect was done by Clark & Martorell (2014) who employed a regression discontinuity approach in which they exploit the fact that a diploma essentially is a piece of paper and that individuals that barely receive it are almost identical to individuals just failing to do so. Their findings suggest that there exists a signaling effect. However, there is far from consensus in existing literature where, for example, Chevalier et al. (2004) find no signaling effect, and Fang (2006) suggest that it accounts for one third of the college wage premium.

The return to schooling has been argued to be heterogeneous, i.e., affecting different groups of individuals differently. For example, Meghir & Palme (2003) showed that increasing the

[2] Economists' failure of addressing selection bias when estimating return to schooling is discussed by Titus (2007) and Porter (2006).

[3] See Card (1999) for a summary over the literature of the relation between education and earnings, focusing on studies using an IV approach.

[4] The signaling and human capital theory are the two most common explanations for the return to schooling, established by Becker (1962) and Spence (1978), respectively. The former states that education is a productivity-enhancing activity which is the reason for higher earnings. Differently put, individuals invest in education to increase their productivity which in turn raises their earnings. According to the latter, individuals signal their productivity level by investing in education as firms have imperfect information of individuals' productivity. Importantly, the cost of signaling correlates negatively with ability making it more worthwhile for high-ability individuals to educate themselves. If the signaling theory holds, it implies that the social return to education could be lower than the private return to schooling which has important policy implications.

compulsory schooling age in Sweden results in a higher return to schooling for high-ability individuals. By using a more extensive data set, Nordin (2008) examines the relationship between ability and return to schooling and finds that when controlling for ability, the return to one year of schooling decreases by almost a quarter. Based on this, Nordin stresses the importance of not attributing the average return to individuals to the bottom and top of the ability distribution. Related to this, it has also been proved that OLS and IV techniques have difficulties in separating realized schooling from ability (Belzil & Hansen 2002).

Propensity score matching approach

To address selection bias and deal with the mentioned problems of OLS and IV, this study uses a propensity score matching (PSM) approach following Rosenbaum & Rubin (1983). The key idea is to find identical individuals that only differ in treatment assignment, that is, have different levels of educational degrees. This enables one to estimate the counterfactual outcome for individuals making it possible to *decompose the treatment effects* on outcomes. This means that differences in initial levels and, more importantly, differences in potential treatment effects are taken into account for all individuals. These facts are not realized when simply applying OLS or IV analysis. For example, OLS and IV controls for characteristics in a linear fashion whereas PSM does not use any given distributions – in other words, matching by propensity score eliminates the linearity assumption.

There exists no convincing natural experiment for educational attainment and it is impossible to randomize it.[5] However, PSM is a quasi-experimental approach that attempts to replicate experimental conditions by correcting for any imbalances, and according to Dehejia & Wahba (2002), a PSM approach yields estimates close to when using an experimental framework. Similar to this study, Titus (2007) uses PSM to find the return to receiving a master's degree in the US and he thoroughly demonstrates how PSM can be utilized to decompose treatment effects. Titus points out the importance of considering the average treatment effect on the treated, and thus not only focusing on the average treatment effect which OLS estimates tend to reflect. He provides evidence that PSM indeed succeeds in doing this.

In contrast to Titus (2007), this study analyses multiple countries. This allows one to find how country-specific variables relate to return to schooling. Also, this study considers the role of individual ability level in the relationship between schooling and earnings. Not controlling for ability would lead to an overstated estimate of the return to school as ability is strongly correlated with educational attainment, which is emphasized by Blackburn & Neumark (1993), Nordin (2008), and Caponi & Plesca (2009).

[5] The closest researchers come to randomisation is by using twins (Ashenfelter et al. 1999; Ashenfelter & Zimmerman 1997) or using instruments for exogenous variations as mentioned earlier.

2. Data

This study relies on data from the Programme for the International Assessment of Adult Competencies (PIAAC) survey from 2016 conducted by OECD. This is cross-sectional microdata consisting of a cognitive test that assesses skills in literacy, numeracy, and reading. This is combined with a rich set of background variables for each individual and most importantly, it includes information about individuals' wage earnings which is the dependent variable in the empirical analysis of this study.

The PIAAC survey is implemented by interviewing approximately 4,000 randomly selected adults from each of the participating countries. The individuals are between 18 and 64 years old and the survey focuses on assessing cognitive and workplace skills by creating indices for literacy, numeracy, and writing skills in technology-rich environments. The survey is designed to be used for cross-country analysis but, however, the accessibility for some variables is not the same for all countries. For example, earnings are not publicly available for all countries, implying that these countries are excluded from this study. Similarly, for some countries, the education level is measured differently from the rest, which also results in exclusion from this study since comparable education levels are crucial. All variables and their characteristics are presented in

Table 1. Summary statistics

	Obs.	Mean	SD	Min	Max
Treatment variables:					
Master $(= 1)$ vs. bachelor $(= 0)$	12,493	.443	.497	0	1
Bachelor $(= 1)$ vs. non-bachelor $(= 0)$	13,739	.506	.5	0	1
Outcome variables:					
Hourly earnings[6]	19,274	20.5	17.5	0	470
Log hourly earnings	19,274	2.67	.990	0	6.15
Covariates:					
Age	19,274	40.9	10.8	18	65
Age^2	19,274	1,793	929	324	4225
Gender $(\text{male} = 0, \text{female} = 1)$	19,274	.445	.497	0	1
Immigrant	19,274	.0751	.264	0	1
Educated in a 'western' country	19,274	.56	.496	0	1
Low educated parents[7]	19,274	.265	.442	0	1
Medium educated parents[8]	19,274	.385	.487	0	1
High educated parents[9]	19,274	.349	.477	0	1
Both parents immigrants	19,274	.0781	.268	0	1
One parent immigrant	19,274	.0492	.216	0	1
Both parents natives	19,274	.873	.333	0	1
Numeracy	19,274	4.34	1.57	0	12.2
Literacy	19,274	4.87	1.27	0	14.4
Writing	19,274	4.57	1.39	0	11.9

[6] Using PPP adjusted dollars.
[7] Low educated parents means that neither parent has attained upper secondary education.
[8] Medium educated parents means that at least one parent has attained secondary and post-secondary, non-tertiary education.
[9] High educated parents means that at least one parent has attained tertiary education.

Table 1. The 22 home country indicators also belong to the covariates but are not shown in the table.[10]

In total, there are almost 20,000 individuals from 22 countries for which there is complete data for all variables. Individuals that are currently in education are excluded and the remaining are divided into three groups based on their *highest* attained educational level: having post-secondary education, bachelor's degree, and master's degree.[11] Individuals that have post-secondary education include both those who have studied at tertiary level and non-tertiary level but most importantly, they have not received a bachelor's degree. Below, Table 2 shows the distribution of individuals over educational attainment. When analyzing the effect of a bachelor's degree, 6,958 individuals are treated and 6,781 are untreated and when analyzing the effect of a master's degree, 5,353 individuals are treated and 6,958 are untreated. When applying the PSM approach, the treated are matched with the untreated.

Table 2. Distribution of educational attainment

Highest attained education level	Observations
Post-secondary education	6,781
Bachelor's degree	6,958
Master's degree	5,535
Total	19,274

Figure 1 presents the percentual differences in means between individuals that have a bachelor's degree vs. those who have post-secondary education but no bachelor's degree. Likewise, Figure 2 presents the percentual differences in means between those with master's degrees vs. those with bachelor's degrees. A difference above zero implies that the higher education level has a higher mean. As expected, there is a significant difference between the

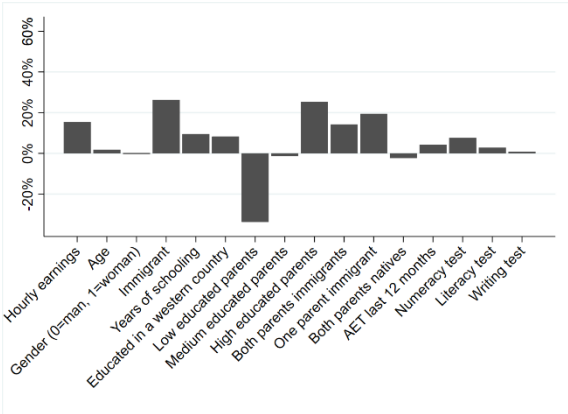

Figure 1. Mean differences between bachelor's degree and post-secondary education

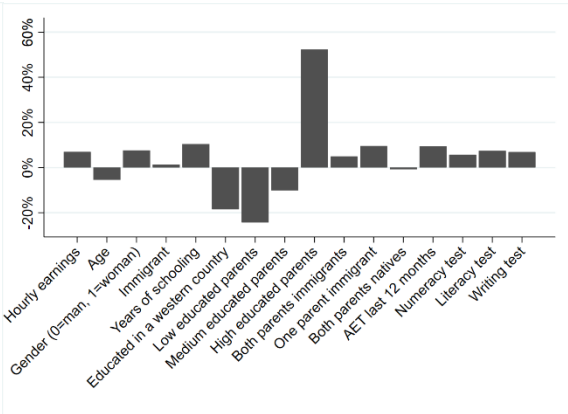

Figure 2. Mean differences between master's and bachelor's degrees

[10] The countries are Belgium, Chile, Czech Republic, Denmark, Ecuador, Finland, France, Greece, Ireland, Israel, Italy, Japan, Kazakhstan, Korea, Lithuania, Mexico, Netherlands, Norway, Poland, Russia, Slovenia, and Spain. The number of individuals from each country varies between 300 and 1,600.
[11] Individuals with no post-secondary education, licentiate or Ph.D. degree are excluded.

groups for most of the variables. Having higher education is for example related to higher earnings, higher educated parents, and higher scores on the ability tests.

3. Method

The propensity score matching (PSM) approach was developed by Rosenbaum and Rubin (1983) and has since then frequently been used when evaluating policies, events, and decisions.[12] The key idea of PSM is to find identical individuals that differ only in treatment assignment and then any difference in the outcome can be attributed to the treatment. There are three main advantages of the PSM approach compared to OLS. First, it makes sure that only individuals that have similar characteristics are compared with each other. Second, it takes into account that individuals with different characteristics react differently to treatment and, thus, allows for heterogeneous treatment effects. Third, unlike OLS, PSM does not rely on the assumption that unobserved variables are normally distributed.

Implementing PSM consists of three steps: estimating the propensity score, choosing the adequate matching algorithm, and estimating the treatment effect. Before clarifying this, two assumptions need to be addressed:

- *The Conditional Independence Assumption* states that treatment assignment is independent of the potential outcomes of the treated and untreated conditional on observable covariates. More formally, it can be expressed as:

$$Y^0, Y^1 \perp\!\!\!\perp D \mid X$$

  where $\perp\!\!\!\perp$ denotes independence. This means that the outcomes Y^0 and Y^1 are independent of treatment D, conditional covariates X. Thus, treatment selection is only based on the observable covariates that are used to determine the probability of treatment.
- *The Common Support Condition*, sometimes called the overlap condition, ensures that there exist treated observations and untreated observations that have similar propensity scores, which enables one to match observations from the two groups with each other. In line with this, all individuals should have a probability between zero and one of being treated. Conversely, the probability of not getting treated should also be between zero and one. This can be written as:

$$0 < P(D = 1|X) < 1.$$

[12] See Caliendo & Kopeinig (2008) for a practical guidance of implement PSM as well as an overview of what fields of study it has been applied to.

Thus, this assumption ensures that there is sufficient overlap between covariates of the treated and untreated making it possible to match these individuals. Fulfilling this assumption is one of the reasons why PSM is preferred over OLS.

Fulfilling these two conditions, Rosenbaum and Rubin (1983) suggest using balancing score $b(X)$ which is a function of the relevant covariates. This solves the dimensionality problem, which means that one can match individuals based on a one-dimensional matching index. The balancing score that is used in this study is propensity score, defined as $\mathrm{p(x) = Pr[D = 1|X]}$, which represents the probability of attaining a certain degree D based on covariates X. To put it differently, if two individuals have the same propensity score, they have the same distribution of covariates and both have a probability between zero and one to get treated, in this case, receiving a degree.

3.1 Estimating the propensity score

The propensity score is estimated by using a discrete choice model, as the treatment variable is binary.[13] The most common discrete choice model used in similar studies is *probit regression* and is, therefore, the main model in this study. The probit model is specified as

$$\mathrm{D} = \beta_0 + \beta_1 \mathrm{X} + \varepsilon$$

where D indicates the binary treatment and X the set of covariates that are relevant for being treated. For convenience, a logit model is also used in one specification, which is the second most common choice model. Identifying and have data of the *relevant* covariates can be problematic when doing PSM as the conditional independence assumption requires that treatment assignment is exogenous conditional on the propensity score, which in turn is based on the covariates.

Choosing covariates

Heckman et al. (1997) and Dehejia & Wahba (1999) argue that including the correct covariates is crucial in order to avoid bias and that the covariates must not be affected by the anticipation of getting treated. Also, the covariates should be unaffected by treatment assignment implying that they cannot be a result of the treatment (Caliendo & Kopeinig 2008). Thus, the covariates should be fixed over time or measured before the treatment and unaffected by the anticipation of treatment. For example, using occupational classification in this study would be incorrect as it is to a high extent a result of educational choices.

It is easier to justify the conditional independence assumption the more informative the variables are but using too many variables can also be problematic for two reasons according to Bryson et al. (2002). First, it can lead to few observations that overlap, i.e. the common support condition may not hold. Second, although irrelevant variables do not necessarily increase the bias

[13] The reason for not using a linear probability model is simply because the estimated probabilities would not be restricted to between 0 and 1.

of the propensity score estimate, they may increase the variance. Heckman et al. (1998: p.1090) suggest choosing the covariates based on statistical significance where one starts with only one covariate, for example, age, and then iteratively adds variables and only keeps the variable if it has a significant relation with the treatment. This approach has been applied in this study, resulting in the covariates presented earlier in Table 1.[14]

3.2 Choice of matching model

After assigning all observations a propensity score, the treated observations are matched with the untreated observations. Several algorithms can be used for this matching procedure and in common, they are all based on the estimated propensity score. As Caliendo & Kopeinig (2008) argue there is no 'winner' of the models that in all cases should be used. Instead, each model has its advantage and disadvantage which can either be amplified or canceled depending on the sample and sample size. Furthermore, Smith (2000) claims that the models are consistent as they asymptotically yield the same estimates as the sample size increases. In common, the choice of model faces a trade-off between bias and efficiency. Based on this, this study uses all models that are presented below, which gives robustness to the interpretation of the results. If the results vary between the models, one should investigate the source of disparity (Caliendo & Kopeinig 2008).

Nearest Neighbour

Nearest neighbour (NN) matching is the most straightforward strategy where each treated individual is matched with an individual with the closest propensity score. NN can be applied in two ways: with or without replacement. With replacement, untreated observations can be matched with more than one treated whereas, without replacement, an untreated observation can only be matched with one treated observation. If matching without replacement is adopted and there are a lot of treated and few untreated individuals with high propensity scores, some of the treated will be matched with low-score untreated individuals. Allowing for replacement overcomes this issue. However, allowing for replacement reduces the number of unique used untreated observations, resulting in higher variance, i.e. there is a trade-off between bias and variance. Three variants of NN matching are employed in this study: NN with replacement, NN without replacement, and NN with replacement using a logit model (for all other specifications, probit is used).

Caliper

Using caliper matching is the same as NN except that the maximum distance to the matching partner is restricted to a certain tolerance level, called caliper. Imposing this eliminates the risk that observations with large propensity score differences get matched making the quality of

[14] Variables that did not pass the statistical significance test is, for example, indicators for if the mother tongue is the same as the main language of the host country and how many years it has been since immigration.

matches higher. The drawback is that fewer observations are used which implies that the variance of the estimates increases.

Radius

A variant to caliper matching is radius matching, in which treated observations are matched with all untreated within the specified caliper. All untreated observations within the caliper are assigned an equally large weight that corresponds to the number of observations within the caliper. This oversampling strategy is suggested by Dehejia and Wahba (2002) and the advantage is that it uses more observations when good matches are available.

Kernel

In Kernel matching, all untreated observations are used for all matches, in contrast to the other described methods where only one or a few are considered. The untreated observations are assigned different weights attached to each treated observation where the weight depends on the difference in the propensity score – the closer each other, the higher weight. As more observations are used the variance becomes lower. According to Frölich (2004), kernel matching produces the most accurate estimates.

3.3 Common support

The next step is to determine the region of common support. This study defines common support according to the *minima and maxima comparison* where all observations with propensity scores lower than the lowest and higher than the highest in the opposite group are excluded. This restriction will barely affect NN with caliper since similar observations are excluded in both approaches. There are two potential issues with the minima and maxima comparison approach. First, it can be problematic if there are observations close to the boundaries but get excluded even though they would be good matches to observations on the other side of the boundary. This is solved by using different specifications like, for example, not restrict radius matching to common support. Second, it fails to identify propensity score regions within the lowest and largest boundaries where no observations are found. This problem is solved when a lot of observations are used with smoothly distributed propensity scores.

3.4 Estimating treatment effects

The average treatment effect (ATE) is the expected effect on the outcome given that treatment is randomly assigned to observations. It is the difference between the expected outcome of the treated and untreated observations:

$$\delta_{ATE} = E(\delta) = E[Y|D = 1] - E[Y|D = 0]$$

where δ is the individual treatment effect. For many policy evaluations, ATE is a relevant effect to consider. However, it includes individuals that perhaps are not intended to be treated and, hence, the estimate reflects something irrelevant. In some sense, it is similar to what OLS is doing

as treated and control observations do not necessarily need to be similar. To put it differently, ATE is "comparing the incomparables", according to Heckman et al. (1997). Instead, it is often more interesting to estimate the average treatment effect on the treated (ATT). This effect reflects what impact treatment has on individuals in the treatment group:

$$E[\delta|D = 1] = E[Y^1 - Y^0|D = 1] = E[Y^1|D = 1] - E[Y^0|D = 1]$$

where Y^1 and Y^0 is the outcome in treatment and control state, respectively, meaning that this estimation requires one to know the outcome for individuals in two different states. $E[Y^0|D = 1]$ is a counterfactual state meaning that it is not observable and needs to be properly substituted to estimate the ATT. This is what PSM aims to solve by matching individuals that are identical to each other except for treatment status implying that selection bias will be addressed.

To find the mentioned effects, I employ the Stata module PSMATCH2 which Leuven and Sianesi (2003) developed. It matches observations based on the propensity scores using NN, caliper, radius, and kernel matching.

4. Results

This chapter is divided into five sections. First, the main results present the return to bachelor's and master's degrees for multiply models across all countries. Second, I assess the matching quality of the main results by performing a balancing test. Third, I estimate the returns separately for each country and relate these estimates to four country-level variables. Fourth, I check whether the returns differ depending on the ability level. Fifth, I combine the country-level analysis with the ability analysis.

4.1 Main results

Figures 3 and 4 illustrate the distribution of propensity scores of receiving a master's degree. The propensity score is estimated by using the covariates presented in Table 1 as well as country dummies. Having a score close to zero implies that the characteristics of the observation are typical for observations without a degree. This could be individuals with, for example, low-educated parents and low test scores. Observations belonging to the green area are matched with observations in the red area. The propensity scores are continuous and the matching is more precise than what the "bins" in the figures indicate meaning that just because an individual belongs to the opposite bin it does not necessarily mean that it is a satisfactory match. When comparing the two figures it becomes clear that when applying NN matching, as in Figure 3, the number of used observations becomes higher compared to when caliper matching is used, shown in Figure 4. The reason is that the only constraint on the number of observations when using NN matching is the common support restriction while applying caliper matching implies a restriction on the quality of the matches.

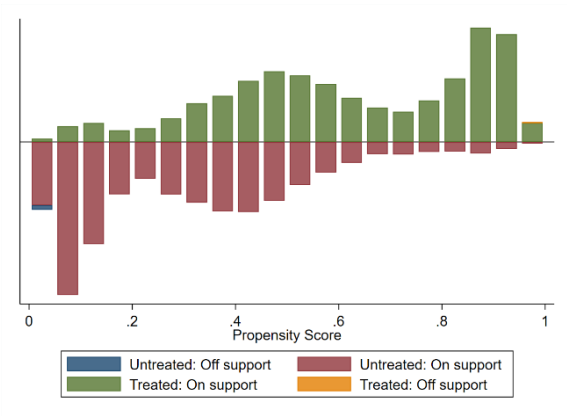

Figure 3. Propensity score histogram of receiving master's degree using NN matching with replacement

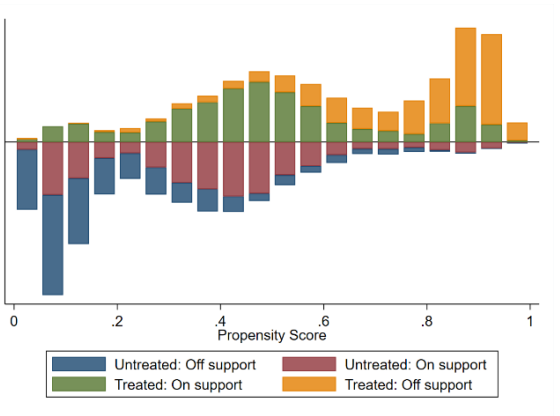

Figure 4. Propensity score histogram of receiving master's degree using caliper

Table 3 shows the main results, i.e. the effects of a bachelor's and master's degree on earnings using eight different models. The dependent variable in all models is defined as the log hourly earnings. As discussed earlier, the most interesting estimate is the average treatment effect on the treated (ATT) which reflects the effect on individuals who receives a degree compared to if they would not have received a degree. Panel A shows that receiving a bachelor's degree results in between 15.0 and 18.3 percent higher earnings according to five of the six PSM models. In panel B, the same five models show that receiving a master's degree is slightly lower: it leads to between 12.5 and 16.4 percent higher earnings than the counterfactuals with a bachelor's degree. These estimates are similar to the average treatment effect (ATE) and the OLS estimates. However, when using NN matching without replacement, in column (3), the estimated effect is only 4.87 percent for bachelor's degree and even negative for master's degree. It is not surprising that when not allowing for replacement the estimates differ from the rest since there is a lot of propensity scores concentrated in the bottom and top of the distribution, as Figures 1 and 2 illustrate. This *forces* observations with low and high propensity scores to be matched with observations in the middle of the distribution. If the propensity score distribution would be smoother, not allowing for replacement would have less impact.[15] Table 4 in Appendix shows the result of the probit selection model, i.e. the predictors of receiving a degree. These results are reported as marginal effects and are used for all PSM models to estimate the propensity scores except for the one in column (2), where instead a logit choice model is used.

[15] Caliendo & Kopeinig (2008) argue that allowing for replacement is preferred when the propensity score distribution is very different between the control and treatment group.

Table 3. Main results

	PSM						OLS	
	(1) NN w. repl.	(2) NN w. repl.	(3) NN w/o repl.	(4) Caliper (0.01%)	(5) Radius (0.01%)	(6) Kernel	(7)	(8)
Panel A. Effect of bachelor's degree on log earnings.								
ATT	.161	.150	.0492	.175	.183	.166		
	(.0274)	(.0274)	(.0168)	(.0250)	(.0232)	(.0218)		
ATE	.164	.155	.0490	.180	.186	.158	.177	.146
	(.0271)	(.0200)	(.0183)	(.0263)	(.0263)	(.0122)	(.0287)	(.0292)
Untreated individuals	6,778	6,778	6,778	4,476	4,476	6,781	6,781	6,781
Treated individuals	6,941	6,942	6,778	4,655	4,655	6,958	6,958	6,958
% of individuals matched/used	99.9	99.9	98.7	66.5	66.5	100	100	100
Panel B. Effect of master's degree on log earnings.								
ATT	.164	.139	-.0490	.159	.141	.125		
	(.0350)	(.0345)	(.0169)	(.0295)	(.0280)	(.0293)		
ATE	.154	.148	.0672	.142	.138	.166	.150	.124
	(.0281)	(.0276)	(.0175)	(.0198)	(.0375)	(.0198)	(.0150)	(.0474)
Untreated individuals	6,920	6,925	5,535	3,520	3,520	6,958	6,958	6,958
Treated individuals	5,535	5,535	5,535	2,931	2,931	5,535	5,535	5,535
% of individuals matched/used	99.7	99.7	88.6	51.6	51.6	100	100	100
Common support	Yes	Yes	Yes	Yes	No	No	-	-
Choice model	Probit	Logit	Probit	Probit	Probit	Probit	-	-

Standard errors in parentheses. The standard errors do not take into account that the propensity score is estimated. When using PSM the standard errors for ATE are bootstrapped. For the OLS estimates, the standard errors are clustered at the country level.

ATT is expected to be lower than ATE as it addresses selection bias in a more appropriate way meaning that ATT to a higher extent controls for heterogeneous treatment effects. However, this pattern only emerges for the effect of a bachelor's degree while for the master's degree it varies depending on the model. This indicates that individuals selecting into bachelor's degrees differ more in characteristics compared to those with post-secondary education than master's degree individuals differ from bachelors. Columns (7) and (8) present the OLS estimates. In (7), the used control variables are the same as the covariates in the PSM models whereas column (8) also includes years of schooling. The latter estimate is lower since years of schooling explain some of the effects.[16]

[16] The reason for why years of schooling is not included when performing PSM is because it makes convergence unachievable, most likely due to high levels of multicollinearity.

4.2 Assessing the matching quality

The matching procedure is based on the propensity score, and thus *only indirectly* based on the covariates. Therefore, it needs to be checked that the procedure balances the distribution of the relevant covariates in both the control and treatment groups (Caliendo & Kopeinig 2008). In other words, the quality of the matching needs to be assessed by making sure that there are no significant differences in the means of the covariates, conditional on the propensity scores. Rosenbaum & Rubin (1983) put it as:

$$X \perp\!\!\!\perp D | P(D = 1|X)$$

i.e. that covariates X are independent of the treatment assignment after conditioning on $P(D = 1|X)$. Differences between the two groups before matching is expected, as shown earlier in Figures 1 and 2, but after matching, the treatment status should be independent of the covariates. Below, in Figure 5, a balancing test is presented that shows the standardized bias for both unmatched and matched individuals. Each "row" represents one covariate and the distance from zero reflects the bias. If the bias is too high for the matched individuals, the matching on the propensity score was not completely successful. The standardized bias for the unmatched, represented by the dots, are on average further from zero than the standardized bias for the matched. Rosenbaum & Rubin (1983) consider biases over 20% as problematic whereas Caliendo & Kopeinig (2008) suggest that it should be below 5%. All matched samples are considerably lower than 20% and only one is larger than 5%. This suggests that the matching procedure is balanced, and the same inferences can be made after performing the same assessment of the other PSM models and for master's degree as well.

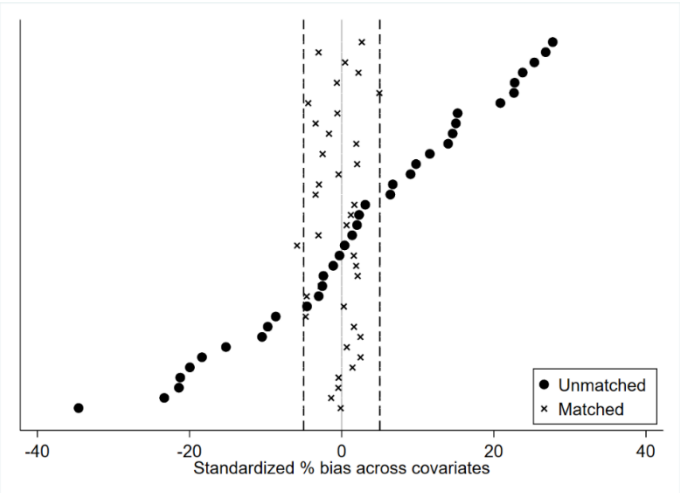

Figure 5. Balancing test

4.3 Country analysis

Table 5 in Appendix presents the effects of master's and bachelor's degrees for each country for which I use NN matching with replacement. Thus, the last row, where all countries are used, corresponds to ATT in column (1) in panels A and B in Table 3. The countries are sorted by the ATT of bachelor's degree.

Although not all estimated returns to schooling for each country are significant, it is still of interest to look at the relationship between the estimates and some country-level variables, especially due to the large cross-country variation of returns. In the scatterplots below, the return to receiving a bachelor's degree is plotted against four different country-level variables. Figure 6 shows that there is barely any correlation between a country's spending on education and its return to receiving a bachelor's degree. However, the relation with social expenditure, which reflects the size of the welfare state, is negative as indicated in Figure 7. The return to schooling seems to be positively related to income inequality, here measured by the Gini coefficient, visualized in Figure 8. More specifically, a 0.1 increase in the Gini coefficient is significantly related with 13.7 percentage points (9.1 percent) higher return to schooling. Finally, Figure 9 indicates that GDP per capita is slightly negatively related to return to schooling as the low GDP countries are spread out. The same patterns emerge when the other PSM models are used and for the return to a master's degree (see Figures 13 to 16 in Appendix).[17]

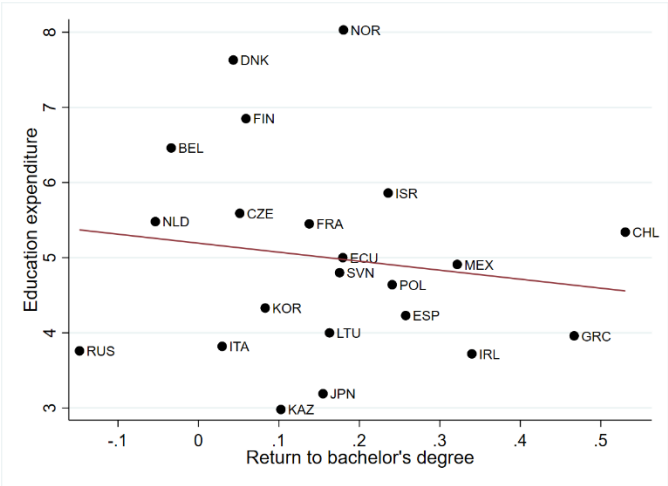

Figure 6. Relation between education expenditure and return to bachelor's degree

[17] Educational expenditure is retrieved from the World Bank and measured as the percentage of GDP. It includes all general governmental spending on all types of educational institutions. Social expenditure is retrieved from OECD:s database and it includes, for example, spending on health care, education, income support programs, and unemployment payments. Consequently, education expenditure is a part of the social expenditure. Data on social expenditure is not available for Ecuador, Kazakhstan, and Russia making them excluded from Figure 4. The Gini coefficient is an index of inequality in equivalized household disposable income taken from the Standardized World Income Inequality Database. GDP is retrieved from the World Bank. All mentioned indicators are from 2016, the same year as the PIAAC survey was conducted.

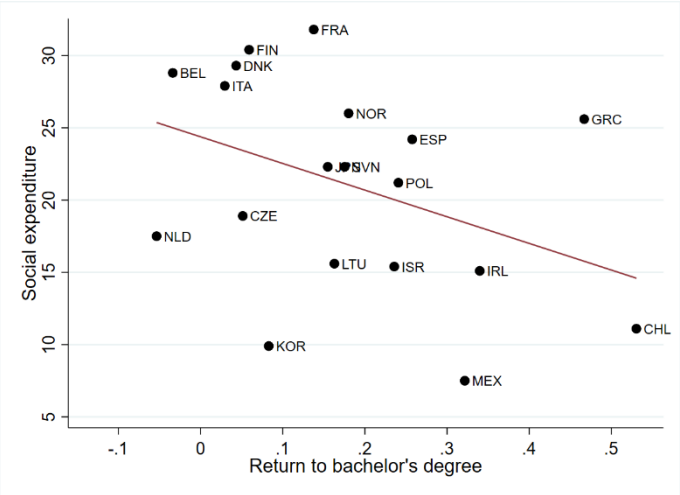

Figure 7. Relation between social expenditure and return to bachelor's degree

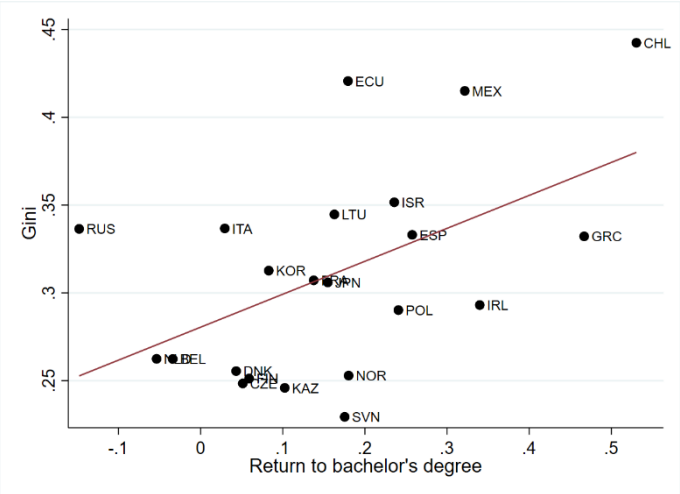

Figure 8. Relation between income inequality and return to bachelor's degree

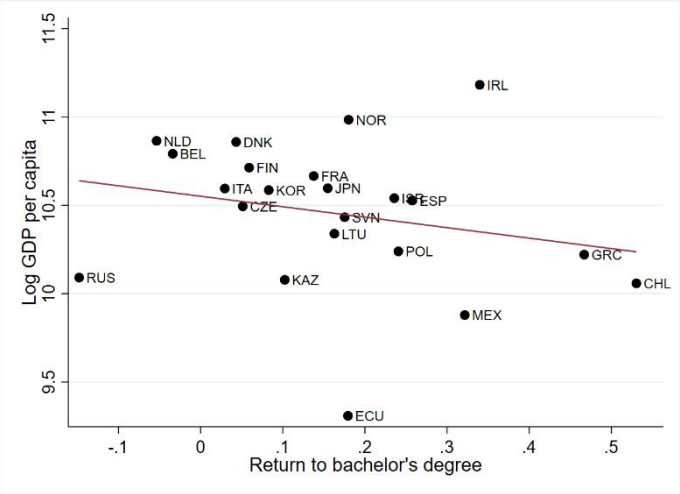

Figure 9. Relation between income and return to bachelor's degree

4.4 Ability analysis

The return to schooling may vary depending on individuals' ability levels. In this section, I examine this by adding the scores from cognitive tests in numeracy, literacy, and writing to construct an ability measurement for each individual. Then, the individuals are divided into ability level quartiles (low, mid-low, mid-high, and high) that are constructed also based on individuals that, for instance, have wrong education level, e.g. only primary education or having a Ph.D. degree. This means that some individuals have an influence on the ability distribution but except for that, are excluded. Consequently, the quartiles are not equally large when analyzing the ability group differences. In fact, the low-ability individuals make up less than a quarter and the high-ability group is larger than a quarter since low educated individuals, which also have lower ability on average, are overrepresented of those who are excluded.

Figures 10 and 11 show the return to a bachelor's and master's degree, respectively, for five of the PSM models that were used in Table 3. *NN without replacement* is excluded in line with Caliendo & Kopeinig's (2008) suggestion to not use this model when the number of observations is low and when there is an uneven propensity score distribution. There is a different pattern between bachelor's and master's degrees. For the former, the return almost has a u-shaped relation with the ability distribution where the mid-high quarter is lower than the rest. For the latter, there is a significant difference between the lowest quarter and the rest.

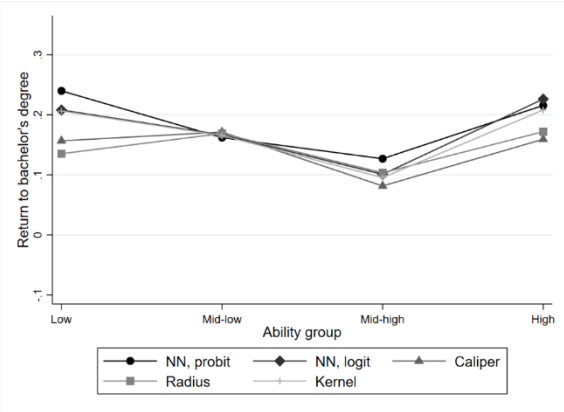

Figure 10. Return to a bachelor's degree across the ability distribution

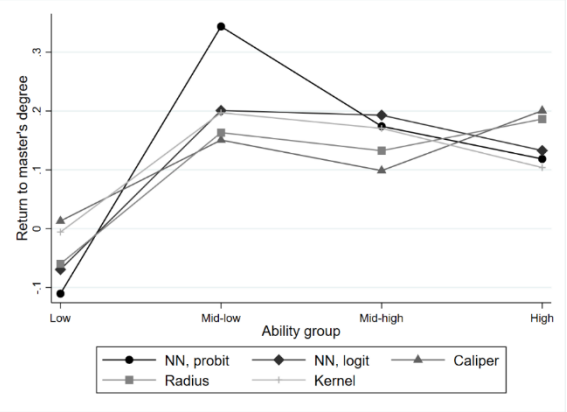

Figure 11. Return to a master's degree across the ability distribution

4.5 Return to schooling, social expenditure, and ability

In this section, I combine the ability analysis with countries' social expenditure. The countries are divided into groups of low respectively high social expenditure, where the cut-off is the median making the groups equally large. Then, the return to a degree is estimated per ability group and country type by taking the average over the five PSM models (NN without replacement is again excluded for the same reason as earlier). The results are presented in Figure 12 where the return to degree represents the average of the return to bachelor's and master's degrees. Low and high

expenditure countries have approximately the same return for the lowest ability group, but for the rest, the returns are strikingly different where the return in low expenditure countries is about 50 percent higher for the two last quarters.

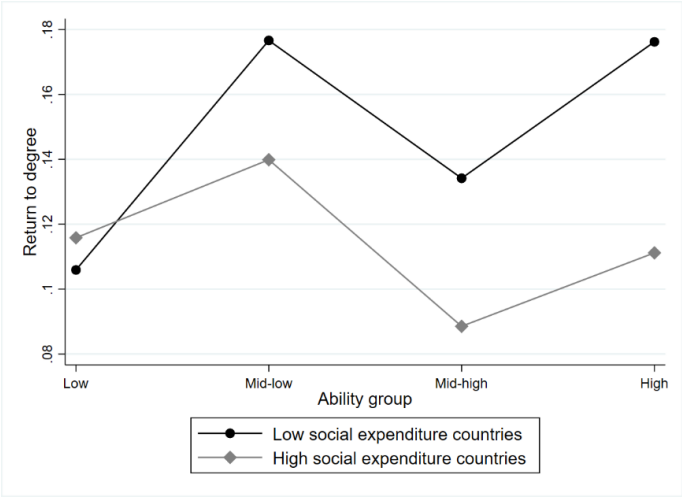

Figure 12. Return to a degree across the ability distribution for low and high social expenditure countries

5. Discussion and Conclusion

Return to schooling is one of the most researched topics in economics due to its complexity and the difficulties in isolating the effect of education. On top of that, different institutional settings between countries and different individual responses to education make the estimation of the return even more complicated. In this study, I have estimated the return in earnings of having a bachelor's and master's degree by applying PSM and, hence, gone beyond the traditional Mincer earnings function. I have disentangled the results by investigating how it relates to country-level variables and individual ability levels. Five interesting inferences can be made.

First, the results suggest that receiving a bachelor's degree results in between 15.0 and 18.3 percent higher earnings than having post-secondary education, and for a master's degree the return is between 12.5 and 16.4 percent higher than having a bachelor's degree. By using PSM the estimated returns are more reliable, compared to OLS and IV estimates, as PSM to a higher extent addresses selection bias. Further, robustness is guaranteed as the estimates are consistent for multiple models backed up with a successful balancing test. The results are similar to previous findings where, for instance, Titus (2007) estimates that a master's degree in Business and Management in the US leads to between 14.7 and 16.3 percent higher earnings. Likewise, Walker & Zhu (2008) estimates the return to a college degree in the UK to approximately 18 percent. But when comparing my results with studies that estimate the return to *years* of schooling, there is a clear difference. Most of the return to schooling years are estimated to just below 10 percent and

some as low as 5% (Trostel et al. 2002). To some extent, it can be explained by the signaling effect since the difference between receiving a degree and completing one additional year of schooling sometimes just is the signaling value of the diploma. To settle this in future research, one needs to incorporate the distinction between years of schooling and receiving a degree, which has been absent in this study.

Second, by using multiple countries I have shown that there is a large cross-country variation of the return to schooling. Countries that have low levels of welfare expenditure, high inequality, and low income are associated with higher returns to schooling. This is very much in line with the results of Trostel et al. (2002) who find similar variations between countries and similar relations to country-specific variables. However, the described relations are merely correlations, and it is difficult to tell what extent, if any, the variables have a causal effect on each other. It still gives interesting indications of what kind of countries that experience low and high returns, but it would be mistaken to suggest that a country should fight inequality by aiming for a lower return to schooling. Instead, one can interpret high returns to schooling as high demand for an educated workforce. In this case, interventions should aim to reduce the cost of higher education, enabling more to pursue higher education.

Third, I have shown that individuals react differently depending on their ability level. For example, the return to a bachelor's degree is twice as large for the high-ability group compared to the mid-high group. Moreover, for the lowest ability quarter, the return to a master's degree is negative while it is between 10 and 20 percent for the other quarters. This is interesting per se, and the pattern of the return to a master's degree is similar to what Nordin (2008) finds. These findings also indicate that low-ability individuals gain from getting a bachelor's degree but when it comes to master's degrees, their counterfactuals, who instead started to work after receiving a bachelor's, earn more. With the same reasoning, the same plot for the return to a Ph.D. degree would perhaps show a negative return for the second quarter as well, i.e., they would benefit from starting to work after receiving a master's degree.

Fourth, the ability measurement in this study is measured *after* the individuals have finished education and potentially gained work experience. Thus, the individuals' education decisions might have affected their ability in this study, which potentially jeopardizes the conditional independence assumption. But if this is the case, finding an increasing return to schooling along the ability distribution supports the human capital theory since it indicates that having higher productivity leads to higher earnings, given that ability reflects productivity. Also, according to the signaling theory, all graduated individuals signal the same productivity level and, thus, should have the same earnings which is not the case for my findings. However, this could be the case for the newly graduated but as one gets older, new work experience enables high-ability individuals to distinguish themselves from others. Since this study uses all adult ages, this explanation cannot be rejected based on my findings.

Fifth, one's ability seems to have a different relation to the return depending on the country's social expenditure. In high social expenditure countries, ability matters less, whereas in low social expenditure countries the return to schooling varies a lot between the lowest and highest ability quarters. In other words, it is more rewarding to have high ability in low social expenditure countries, where the returns are 50 percent higher than in high social expenditure countries. One interpretation is that countries with an extensive welfare state (i.e., high social expenditure) succeed in having a more equal return to the same level of education. Generous pensions, social insurances, and unemployment benefits, but also efforts to have equal payments to equal jobs, could be explanations for this pattern. Alternatively, it could be the case that the signaling effect is larger in high social expenditure countries (your salary is highly based on your education level) whereas the effect from human capital accumulation is more decisive in low social expenditure countries (your ability is reflected in your salary).

In conclusion, the findings of this study have been multiple and relevant for different areas in the sense that both country- and individual-level characteristics have been considered. I have provided evidence showing that the return to schooling differs between different types of countries and varies across the individual ability distribution. However, the breadth has to some extent been at the expense of the accuracy of the discoveries. To improve this, the most important modification would be to use panel data instead of cross-sectional. This enables one to find changes over time, how the returns differ between cohorts, and, most importantly, how changes in country-level variables relate to the return which would give a deeper understanding of the causal relationship.

6. References

Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?. The Quarterly Journal of Economics, 106(4), 979-1014.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. Labour economics, 6(4), 453-470.

Ashenfelter, O., & Zimmerman, D. J. (1997). Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers. Review of Economics and Statistics, 79(1), 1-9.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of political economy, 70(5, Part 2), 9-49.

Bedard, K. (2001). Human capital versus signaling models: university access and high school dropouts. Journal of Political Economy, 109(4), 749-775.

Belzil, C., & Hansen, J. (2002). Unobserved ability and the return to schooling. Econometrica, 70(5), 2075-2091.

Blackburn, M. L., & Neumark, D. (1993). Omitted-ability bias and the increase in the return to schooling. Journal of labor economics, 11(3), 521-544.

Bryson, A., Dorsett, R., & Purdon, S. (2002). The use of propensity score matching in the evaluation of active labour market policies.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.

Caponi, V., & Plesca, M. (2009). Post-secondary education in Canada: can ability bias explain the earnings gap between college and university graduates?. Canadian Journal of Economics/Revue canadienne d'économique, 42(3), 1100-1131.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. NBER working paper, (w4483).

Card, D. (1999). The causal effect of education on earnings. Handbook of labor economics, 3, 1801-1863.

Chevalier, A., Harmon, C., Walker, I., & Zhu, Y. (2004). Does education raise productivity, or just reflect it?. The Economic Journal, 114(499), F499-F517.

Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. Journal of Political Economy, 122(2), 282-318.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American statistical Association, 94(448), 1053-1062.

Fang, H. (2006). Disentangling the college wage premium: Estimating a model with endogenous education choices. International Economic Review, 47(4), 1151-1185.

Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. Review of Economics and Statistics, 86(1), 77-90.

Harmon, C., & Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. The American Economic Review, 85(5), 1278-1286.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. The review of economic studies, 64(4), 605-654.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. Econometrica, 1017-1098.

Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. Econometrica, 73(3), 669-738.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. Handbook of the Economics of Education, 1, 307-458.

Kane, T. J., & Rouse, C. E. (1993). Labor market returns to two-and four-year colleges: is a credit a credit and do degrees matter? (No. w4268). National Bureau of Economic Research.

Leigh, A., & Ryan, C. (2008). Estimating returns to education using different natural experiment techniques. Economics of Education Review, 27(2), 149-160.

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.

Maluccio, J. A. (1998). Endogeneity of schooling in the wage function: Evidence from the rural Philippines (No. 583-2016-39673, pp. 1-58).

Meghir, C., & Palme, M. (2003). Ability, parental background and education policy: empirical evidence from a social experiment (No. 03/05). IFS Working Papers.

Mincer, J. (1974). Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

Montenegro, C. E., & Patrinos, H. A. (2014). Comparable estimates of returns to schooling around the world. The World Bank.

Nordin, M. (2008). Ability and rates of return to schooling—making use of the Swedish enlistment battery test. Journal of Population Economics, 21(3), 703-717.

Porter, S. R. (2006). Institutional structures and student engagement. Research in Higher Education, 47(5), 521-558.

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. Education Economics, 26(5), 445-458.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Smith, J. (2000). A critical survey of empirical methods for evaluating active labor market policies (No. 2000-6). Research Report.

Spence, M. (1978). Job market signaling. In Uncertainty in economics (pp. 281-306). Academic Press.

Titus, M. A. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: An application to the private returns to a master's degree. Research in Higher Education, 48(4), 487-521.

Trostel, P., Walker, I., & Woolley, P. (2002). Estimates of the economic return to schooling for 28 countries. Labour economics, 9(1), 1-16.

Walker, I., & Zhu, Y. (2008). The college wage premium and the expansion of higher education in the UK. Scandinavian Journal of Economics, 110(4), 695-709.

7. Appendix

Table 4. Predictors of receiving a degree using probit model

	Marginal effect on bachelor's degree	Marginal effect on master's degree
Belgium	Reference	
Chile	1.11	-2.91
Czech Republic	.765	-.499
Denmark	.836	-.968
Ecuador	2.11	-3.04
Spain	1.85	-1.17
Finland	1.49	-1.32
France	1.61	-1.30
Greece	1.71	-1.70
Ireland	1.29	-1.58
Israel	1.95	-1.86
Italy	3.19	-2.44
Japan	1.42	-2.90
Kazakhstan	2.09	-3.26
Korea	1.45	-2.84
Lithuania	1.35	-1.68
Mexico	2.61	-3.00
Netherlands	2.50	-1.78
Norway	1.95	-1.63
Poland	1.46	-.559
Russia	.363	-.0639
Slovenia	1.21	-2.79
Age	-.0238	.0633
Age^2	.000191	-.000634
Female	.00925	.0449
Immigrant	.148	-.0738
Educated in western country	-.368	-.340
Low educated parents	-.482	-.374
Medium educated parents	-.333	-.312
High educated parents	Reference	
Both parents immigrants	.00730	.149
One parent immigrant	-.0484	.0855
Both parents natives	Reference	
Numeracy	.0166	.0408
Literacy	.116	.0756
Writing	.0268	-.00705
Constant	-1.28	-.291

Table 5. ATT on hourly earnings per country using NN with replacement

	Bachelor's degree			Master's degree		
Country	Obs.	% matched	ATT	Obs.	% matched	ATT
Chile	436	99	.530 (.117)	233	87	.288 (.173)
Greece	340	96	.467 (.167)	231	90	.420 (.210)
Ireland	1,059	99	.340 (.0994)	665	98	.0298 (.106)
Mexico	253	93	.322 (.313)	255	88	.121 (.363)
Spain	409	97	.258 (.141)	486	98	.303 (.119)
Poland	239	96	.241 (.139)	768	99	.204 (.106)
Israel	574	98	.236 (.103)	559	97	.266 (.102)
Norway	1,044	98	.180 (.0659)	1,090	99	.0878 (.0500)
Ecuador	216	88	.179 (.149)	193	90	.215 (.265)
Slovenia	569	98	.175 (.0931)	325	91	-.158 (.169)
Lithuania	640	97	.163 (.0986)	651	99	-.00888 (.0983)
Japan	1,202	99	.155 (.0769)	759	95	.246 (.117)
France	732	98	.138 (.0778)	659	99	.252 (.0878)
Kazakhstan	791	97	.102 (.0807)	616	61	.0390 (.227)
Korea	1,056	99	.0829 (.0746)	695	92	.186 (.201)
Finland	1,043	99	.0590 (.0650)	862	98	.268 (.076)
Czech Republic	190	93	.0513 (.159)	415	89	-.139 (.141)
Denmark	1,168	99	.0433 (.0738)	798	99	.0832 (.0842)
Italy	222	95	.0294 (.240)	264	87	.186 (.199)
Belgium	605	94	-.0339 (.136)	354	86	.102 (.0829)
Netherlands	574	99	-.0536 (.117)	693	97	.276 (.0891)
Russia	250	74	-.149 (.249)	440	78	.104 (.246)
All countries	13,739	99.9	.161 (.0274)	12,493	99.7	.164 (.0350)

Standard errors in parentheses. The standard errors do not take into account that the propensity score is estimated.

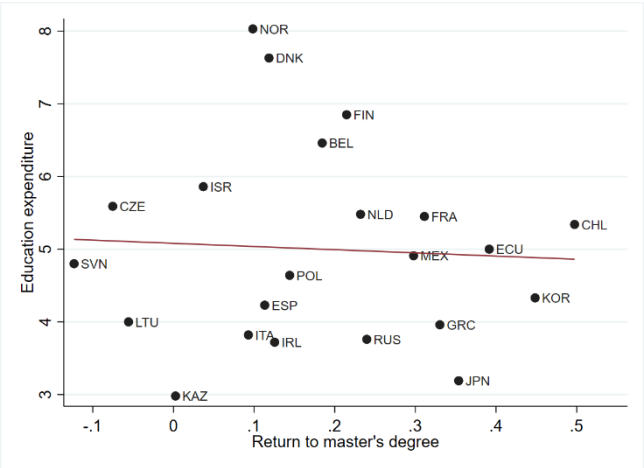

Figure 13. Relation between education expenditure and return to master's degree

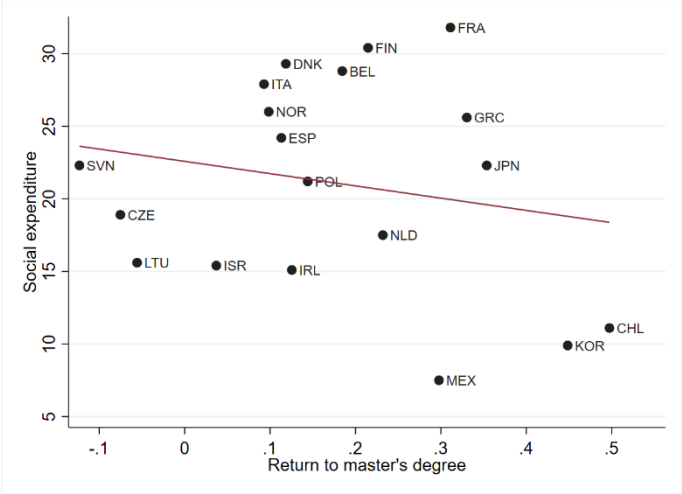

Figure 14. Relation between social expenditure and return to master's degree

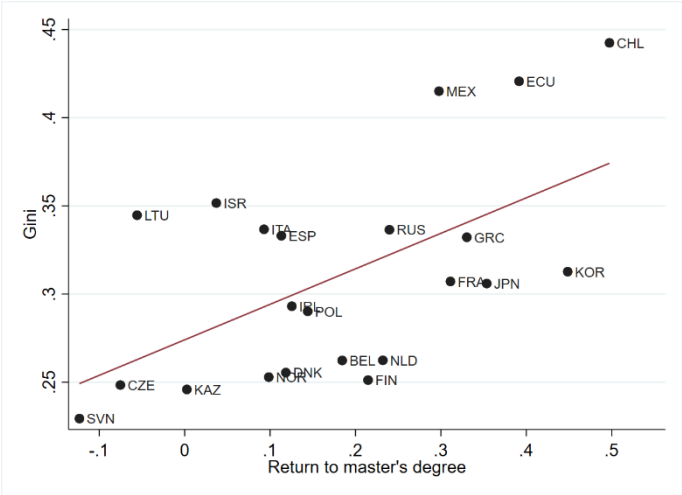

Figure 15. Relation between income inequality and return to master's degree

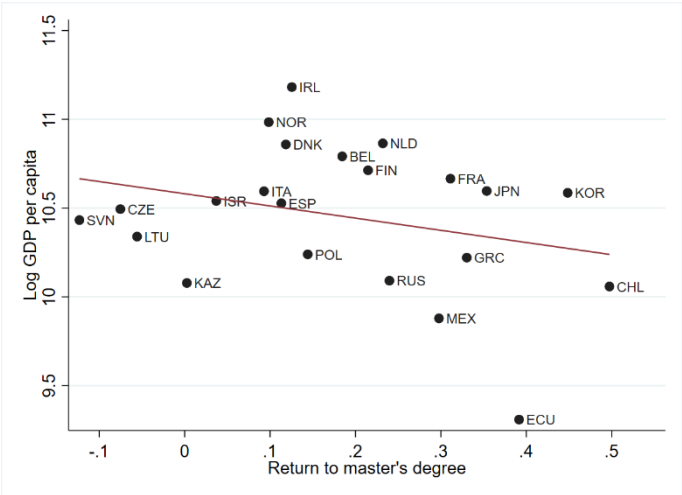

Figure 16. Relation between income and return to master's degree