Returns to Education Attainment:

Eugene Trostin *University of Ottawa*

Abstract. An individual's education attainment is one of the major determinants in their future employment decision - certifying a higher wage. Using the *National Longitudinal Survey of Youth | 1997-2017* (NLSY97) data, I examine matched individuals who have the same level of education attainment and observe key characteristics, such as parental household income, to explain differences in their wages in 2010. When using propensity score matching and kernel matching the data suggests that wage disparities between genders become smaller as individuals attain higher degrees. OLS estimates reveal statistically significant results for returns to education attainment. Using controls such as parental income status, race, and gender allow to identify associations between wage disparities.

JEL Classification: I23, I26, J16

1. Introduction

Returns to education attainment "is a long standing question in Economics" (Mohen). The dynamic nature of people, markets, and policies tend to change over time warranting attention from politicians and researchers. Using characteristic data, this analysis explains differences in observed wages explained by the social origins hypothesis and the ascriptive hypothesis despite the individual's education attainment.

1.1 Subject of the Paper:

Education is defined as the process of teaching the necessary knowledge deemed important by society such as literacy, mathematics, and socialization skills. The pursuit of education is best explained with the societal benefits that it brings and the human capital investment which makes workers more efficient. Individuals who have not completed primary education are earning lower wages in comparison (see figure 1 in appendix). The option of further specialization can be explored by continuing schooling with post-secondary education or training certifications. Having a society that has more education makes everyone better off in terms of human development which is captured by an education index[1]. With more individuals who possess higher education managers tend to pick and choose between good candidates and those who they would prefer to work with[2]. For the same level of education variance in wages can be explained between individuals with their ascribed status[3] and individuals social origins[4].

[1] (UNDP 2020). One dimension of the HDI calculation is education. It is calculated by taking the average between expected years of schooling and mean years of schooling. The HDI rank can be located on table 2 page 351.

[2] "Thousands of randomly manipulated resumes were sent in response to online job postings in Toronto to investigate why immigrants, based on skill, struggle in the labor market. The study finds substantial discrimination across a variety of occupations towards applicants with foreign experience or those with Indian, Pakistani, Chinese, and Greek names compared with English names" (Oreopoulos 2011, pg. 148).

[3] "Ascription Hypothesis: Wage attainment reflects otherwise unexplained, discriminatory factors pertaining to gender and race/ethnicity or nativity that operate above and beyond school effects and social origins" (Elman and O'Rand 2004, 127)

[4] "Social Origin Hypothesis: Adult wages reflect relative childhood advantage … that is positively associated with parental socioeconomic status and putative individual ability" (Elman and O'Rand 2004, 128).

1.2 Subject of Interest

Identifying differences in wages despite similar education is a one major concern for individuals joining the labour force. The expensive investment in post-secondary education should be remunerated for their investment and time spent in specializing in their field. Although observed differences in geography, ascriptives, social origins, and skill help explain wage differentials for individuals who have the same level of education.

2. Literature Review

2.1 Economic Theory

Education is an investment decision made by an individual. After completing mandatory schooling individuals have a choice of either continuing their education or taking a career option instead. After completing their education the individual will likely work in a job unrelated to their education, the pursuit is to eventually be hired in their field to earn a higher wage[5] from their specialization. On the other hand an individual can choose a career option and be earning wages right away, but compared to an educated (specialized) individual who is working in their field of study, would earn a lower wage.

2.2 Relevant Literature

The Race Is to the Swift: Socioeconomic Origins, Adult Education, and Wage Attainment[6] is a sociological article that is beneficial in two major aspects. (1) They provide definitions outlining key theories like ascription hypothesis and social origins hypothesis. (2) The other aspect is them finding differences in returns to education depending on the individual's

---

[5] (Chapter 5 Bhattacharya, Hyde and Tu 2014)
[6] (Elman and O'Rand 2004 123-126)

socioeconomic origins. The argument is that wealthier and stable households have an increased likelihood of their kids continuing schooling.

Estimates of the Economic Return to Schooling from a New Sample of Twins[7] is a survey which took data from identical twins. They find a perfect counterfactual by surveying monozygotic twins. They are perfect in terms of internal validity, because they are not only genetically identical, but they also go to the same school and live with the same parents. In essence controlling for nature in the Nature vs. Nurture debate. Although one main criticism of this study is in the way they obtained their survey data. They set up a booth in a Twins Day Festival in Ohio (1991) and advertised that if the twins would participate in the study they would get a scholarship upon completion of the survey. This methodology may have led to an overestimation of their results because the individuals voluntarily self selected themselves into this study.

2.3 NLSY97 Literature

Differentiation and Work: Inequality in Degree Attainment in U.S. Higher Education[8] this study uses the NLSY97 data set from 2009 and argues that a "more complex set of patterns is revealed when examining the relationship between employment, family background, and degree attainment across different institutional types and educational credentials" (Roksa 293). The main conclusion is that there is a lower likelihood of poorer households continuing their education opting to either focus more on work or selecting cheaper options like community college (two year programs that emphasise training).

---

[7] (Krueger and Ashenfelter 1992)
[8] (Roksa 2011)

The Changing Benefits of Early Work Experience [9] This study compares two cohorts from 1987-1989 (NLSY79) and 2008-2010 (NLSY97). Baum and Ruhm stress that "high school work experience has changed over the last 20 years", students who work in their senior year experienced a diminished average annual earnings premium by approximately 5% compared to the 1987-1989 cohort.

Two major issues with this study is that there are major differences between the two cohorts that would lead to unrepresentativeness. NLSY97 surveys were administered online through Computer-Assisted Personal Interviewing (CAPI). For NLSY79, these surveys were conducted with an interviewer asking the respondent the survey questions[10]. This may lead to respondent fatigue as the interviewer would ask questions that would span 104 pages. Baum and Ruhm chose the years 1987-1989 and 2008-2010. The great recession in 2008 would have led to them to underestimate their results.

The Changing Roles of Education and Ability in Wage Determination[11] They argue "that changing technology has led to reforms in the schooling system, which has resulted in a more relevant and merit-oriented education" (687). In essence that between 1980 and 2000 technology changed so much that it impacted the labour market. Somewhat similar to self-checkout terminals impacting labour markets changing the dynamic for unskilled labour.

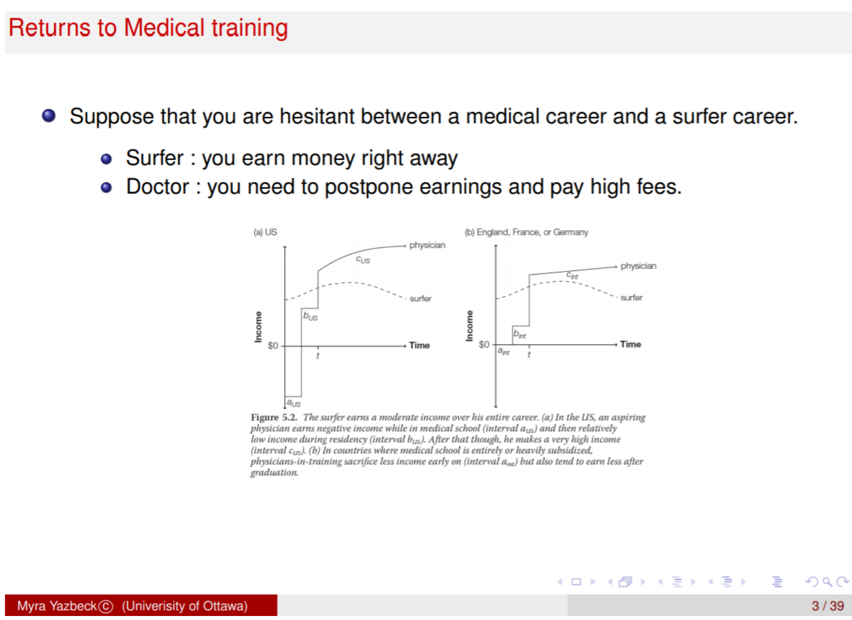
[9] (Baum and Ruhm 2016)
[10] (NORC 1987 pg. 5)
[11] (Castex and Dechter 2014)

3. Model

3.1 Economic Theory

There are four aspects to keep in mind in terms of economic theory. (1) Human Capital, (2) Returns to education/training,(3) Signaling theory, (4) and how technology impacts the labour market. This subsection deals with their definitions.

1. Human Capital Theory suggests that education can nurture those who can contribute to the country's expansion (higher GDP due to more efficient workers).
2. An individual who chooses to continue their education forgoes the career option, as a consequence they "postpone earnings and pay higher fees [tuition]"[12]. Labour economics would use net present values, discount factors, internal rates of returns to calculate which is an optimal choice for the individual - treating education as an investment decision. For example an individual who cannot discount the present will choose the career options instead as they would need the money now.



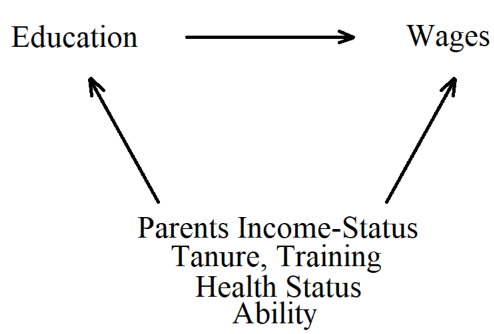
Returns to Medical Training/Education[12]

[12] Yazbeck (2019)

3. Michael Spence signaling theory argues that hiring investment decisions are done under uncertainty[13]. Potential hires can signal to managers and compete with other resumes by filling in the information asymmetry. Spence writes that “signaling power of education, job experience, race, sex, and a host of other observable, like personal characteristics” (355) are used to help managers to pick between candidates.
4. Castex and Dechter research emphasizes that wage determination over the last 20 years (1980-2000) has been impacted with a large shift in technology. This can be extended to 2000-2020 as technology has been shifting some industries. My model does not take technology into account but specific industries like the service industry have found ways to optimize workers. An example of this would be self checkout lowering the amount of workers on registers focusing labour elsewhere.

3.2 Estimation Model

The empirical model identifies how education impacts an individual's wages, there is a case to be made in regards to omitted variable bias which I need to control for.



Model specification - Omitted Variable

[13] (Spence 1973)

The NLSY97 survey asked participants to give the survey to their parents and fill in the gross households income. To simplify I have taken this data and created a categorical variable for income status using definitions provided by the PEW research centre.

Income Status:	Category	Income
Low-Income	(1)	< 31000$
Low-Middle Income	(2)	31000-42000
Middle Income	(3)	42000-126000
Upper-Middle Income	(4)	126000-188000
High Income	(5)	>18800

NLSY97 parental income status categorical variable

Tenure and training were asked in a yes or no manner while health status and ability were asked in a categorical manner where the respondent had to rank it with 1 (low) to 5 (high). I would argue for a potential case of response bias for health status as it asked "in general, how is your health?". This may have led respondents to select option five more often as the question is vague. One aspect I do not control for is the household environment, namely things like if the parents are divorced and live separately, parenting style, if children were adopted, if there were more children in the household, and which kind of education degree they have attained.

$$\underbrace{Y_i}_{\text{Wages}} = \beta_0 + \beta_1 \underbrace{X_i}_{\text{Education}} + u_i$$

Baseline specification

Normally this regression model would be converted into log wages but in this case it would be best to leave it as it is. Wages in 2010 follow a normal distribution that is skewed right (figure 2 in appendix), if converted into log wages the distribution has a larger left skew (figure 3 in appendix). To make interpretation easier to see with graphs I elected to leave wages as is.

Taking a similar approach to health sciences, it would be possible to match. Like this case study (Littnerová, Jarkovský, Pařenica, Pavlík, Špinar and Dušek 2013). Matching forgoes generalizability but makes it possible to tease out a causal inference, although it imposes the need to provide a sub-sample with common support.

King and Nielson recommend not to use propensity score matching, recommending to use kernel matching instead as this algorithm is more efficient. In terms of assumptions: unconfoundedness is plausible, but I did provide the sufficient support to claim with certainty that there is, Heckman Selection Technique will help this problem as it will randomize observational data strengthening assumption[14]. In terms of common support I use code introduced in this seminar[15]. I match the sample on variables such as race and ethnicity, parental income class, urban rural, general health, personality scale (ps1-ps8)[16] and marital status.

3.3 NLSY97 Data

I provide summary statistics for the data used in this research (figure 4 in appendix). I would like to bring attention to a few aspects. Race and ethnicity lacks a lot of detail as it only accounts for African Americans, Hispanics, Mixed, and Caucasians. Due to security reasons geocodes are not provided by the NLSY97 and only give region data which takes fifty states and separates them by four regions (north, south, east, west). This brings out the issue of differences between within states policies and minimum wages.

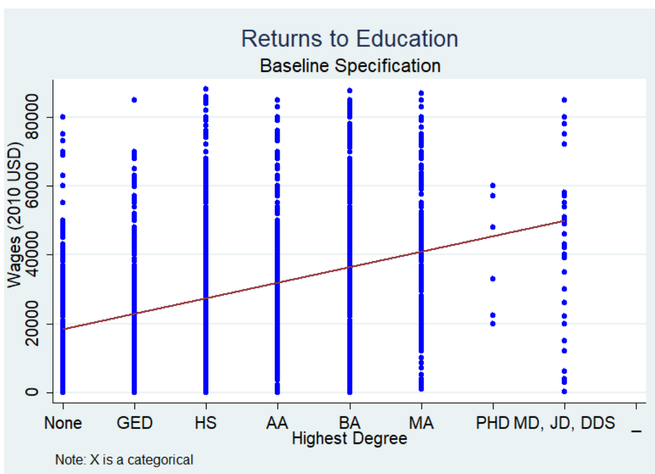
[14] ("Stata Heckman Selection Model - Manuel," 2020)
[15] (Jann 2017) A seminar providing the code and explanations for kernel matching
[16] eg. PS1= Hardworker, PS8= Bending Rules

4. Presentation of Results

4.1 OLS

The data set comes with sample weights to make the data more representative of the general US population. With that in mind this methodology unfortunately is to weak to provide a causal interpretation but it allow for generalizability as a whole [17]. Please visit this link to access the output table for the regression analysis.

For the baseline specification I found that, an increase in education by one degree is associated with, on average, an increase in total income in wages by $4,243. This result is statistically significant at the 1% level.



Graph 1 - Baseline Specification

In regression seven, an increase in education by one degree is associated with, on average, an increase in total income in wages by $4,084 controlling for gender, race, marital

[17] Rundel M.C (2013) see figure 5 in appendix. The data was randomly sampled but there is no random assignment

status, region, union, training, and parental income status. This result is statistically significant at the 1% level.

In regression ten, I've interacted education onto itself to show the subcategories averages. This was also done for marital status, race, and parental income status. For education compared to individuals with no degrees all results show that achieving a higher degree is associated with a higher wage. All results for interacted education are statistically significant at 1%, the difference between no degree and other degrees grows in magnitude with the mode difficult degrees .

Discrimination in regards to gender and race does persist showing when comparing to African Americans to Hispanics ($8137), and mixed race ($4450) to Caucassian ($8874). Hispanics and Causassian averages are statistically significant at 1% level. In regression three, there is an average difference in wages by -7288$ for women.

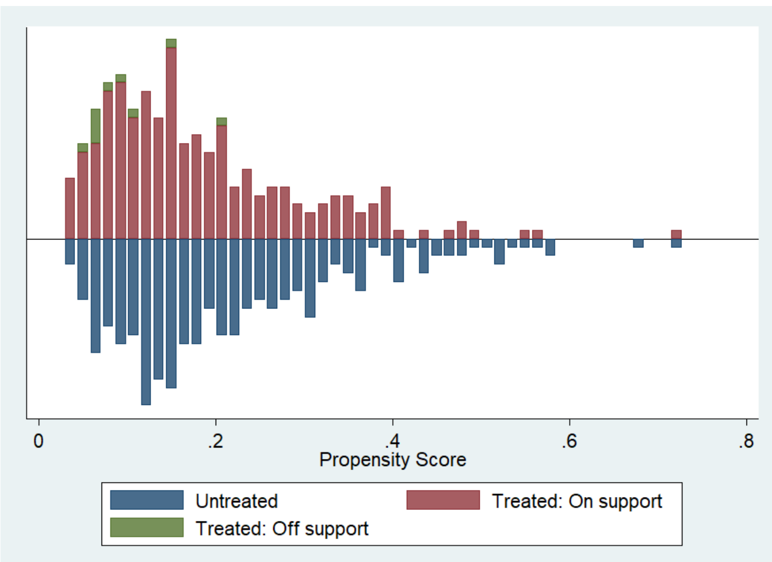
4.2 Matching

Two different matches algorithms were used to obtain causal inference[18], the matching is done with individuals who have similar personality scores (PS1-PS8), education, occupation, industry, marital status, race, and parental income status. Using propensity score matching I find that characteristics gender, and race create differences in wages despite having the same level of education. There is a slight concern with efficiency in using propensity score matching[19], for the sake of comparison a kernel match will be done as well.

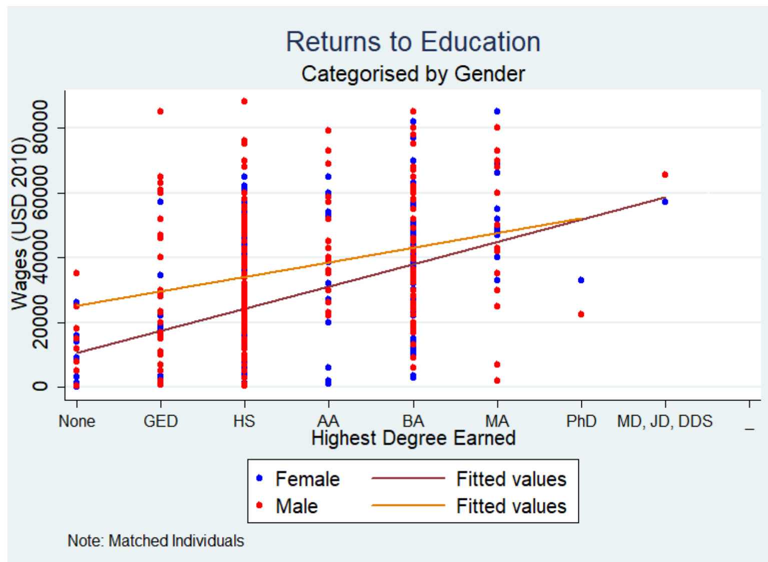
---

[18] Rundel M.C (2013) this approach forgoes generalizability to the US population that the OLS had but permits a causal conclusion, albeit only for the sample.
[19] (King and Nielson 2019)

Starting with propensity score matching, there is a case for common support. When matching for gender the regression appears to suggest that as individuals pursue higher education the wage disparity becomes smaller and converges. Although it is worth noting that there is a smaller amount of observations for PhD and other professional degrees (MD, JD, DDS) which may have affected the matching algorithm. The treatment effects even when matched education there are still differences in wages even when matching on



Graph 2 - PSM - Common support

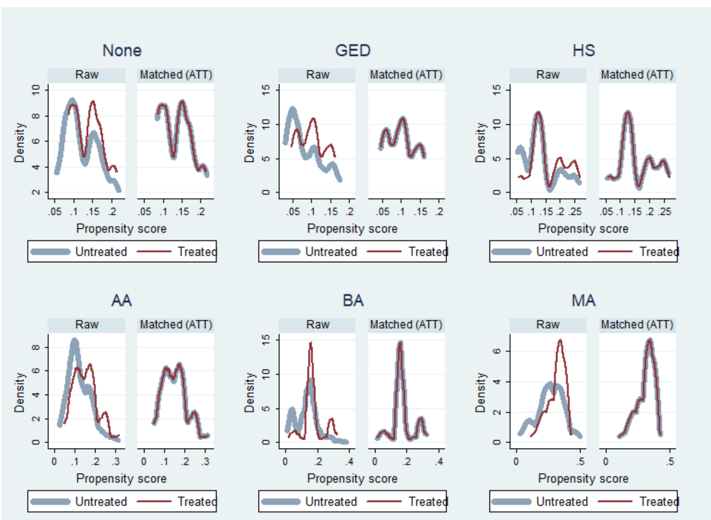


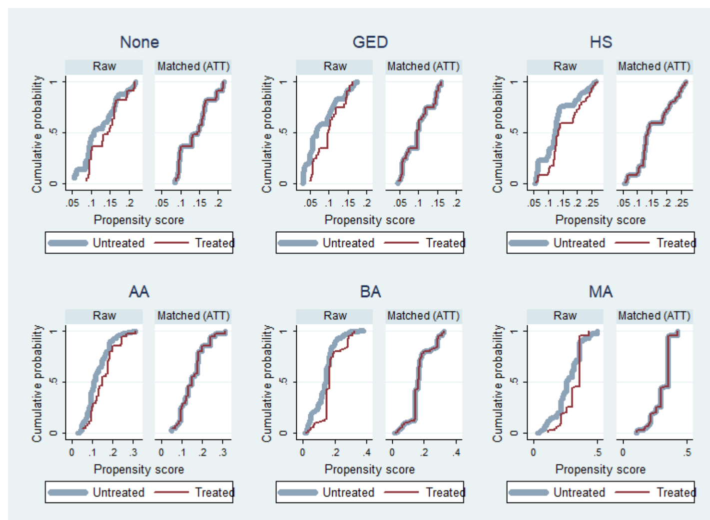
Graph 3 - Returns to Education by gender

To see the regression table for matched individuals please visit this link. For the baseline specification I found that, an increase in education by one degree is associated with, on average, an increase in total income in wages by $5,640. This result is statistically significant at the 1% level. This result is a lot higher when compared to OLS but it is worth noting two reasons for this higher wage. The amount of observations was reduced to 500 (256 for treatment and 260 for control). The statistical weights were also removed as they are no longer representative of the US population.

In regression three, an increase in education by one degree is associated with, on average, an increase in total income in wages by $5,036 controlling for gender, urban, training program, race, union, and parental income. This result is statistically significant at the 1% level. For gender the wage difference is -$5,926 and it is statistically significant at the 1% level. For race when comparing to African Americans there is an increase in wages for Hispanics by $2,100.

Following Jann tutorial on kernel matching, there is an evident case for common support with the variables specified in section 3.2.





Graph 5 - Kmatch density common support

Graph 6 -Kmatch cumulative probability

Kernel matching has identical results to propensity score matching. Although due to collinearity I had to remove personality scores and training programs. A convenient future with kernel matching is that it is able to show the Average Treatment Effects by subcategory. Based on gender individuals with higher degrees the wage differences starts to converge (see figure 6). Women with PhD and professional degrees surpass male wages, but there are caveats to these findings. For PhD there is only one control and for professionals degrees there are twenty females and fourteen males. There are simply not enough individuals in this matched sub-sample. It would be convenient to have more observations for PhD and professional degrees to provide a more confident conclusion, that being said it is possible to re-do this experiment with NLSY97 datasets after 2010 but that only reveals if the conclusion can be reproduced.

4.3 Implications

The main implications are difficult to discern as this level of observation does not account for geographical data nor which degree type of degree they attained. With OLS it can be generalized that education degrees yield higher wages on average. It can also be said that depending on characteristics like gender (see figure 7), race (see figure 8), or parental income status (see figure 9) the wages start to differ. When matched to individuals with similar levels of education there are differences in wages between genders, although as higher degrees are attained the two slopes start to converge. For kernel matching ATE surpasses male counterparts for PhD and professional degrees.

The main implication appears to be that this sample experiences differences in wages and its association with their gender, even though they possess the same level of education furthering the credibility towards the ascriptive hypothesis.

5. Conclusion

In this paper I analyze returns to education for the National Longitudinal Study of Youth (NLSY97) and discover that despite having the same level of education, women on average earn lower wages in America. This difference becomes smaller for the observation who possess higher level degrees. The data suggests that higher education individuals are highly favoured earning a higher wage on average. Using OLS and statistical weights these findings can be generalized to the rest of the US population. Unfortunately due to limitations with geocode access prevented the ability to control for state level data, meaning aspects like higher minimum wage in other states or specific state level policies are not accounted for. With statistical matching both algorithms (propensity score matching and kernel matching) reveal wage differentials between men and women who have the same level of education. These differences start to grow smaller as education attainment rises. One main limitation of this study is that unconfoundedness is plausible, but not certain. There are techniques which can further strengthen this assumption [20] although it is not possible to prove. This quasi-experiment reveals that education is associated with higher wages and that wage disparities between genders become smaller as higher levels of education are achieved.

[20] (Tucker, Thank, Aitsahlia, Demski, Donohoe, Flannery, Greene, Kirk, Omer, Schroeder and Tse 2011) provides ways to tackle selection biases

References

Baum, C.L., and C.J. Ruhm (2016) "The Changing Benefits of Early Work Experience.," Southern Economic Association October. available at. Last updated October 29, 2020

Bhattacharya, J., T. Hyde and P. Tu (2014) *Health Economics*. Basingstoke: Palgrave Macmillan

Blundell, R., L. Dearden and B. Sianesi (2005) "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey," *Journal of the Royal Statistical Society: Series a (Statistics in Society)* 168, 473–512 doi:10.1111/j.1467-985x.2004.00360.x

Castex, G., and E.K. Dechter (2014) "The Changing Roles of Education and Ability in Wage Determination," October. available at . Last updated October 29, 2020

C.M. Achilles, Helen Pate Bain, F. Bellott, J. Boyd-Zaharias, J. Finn, J. Folger, J. Johnston and E. Word (2008) "Tennessee's Student Teacher Achievement Ratio (STAR) Project," *Harvard Dataverse*, October doi:hdl:1902.1/10766

Elman, C., and A.M. O'Rand (2004) "The Race Is to the Swift: Socioeconomic Origins, Adult Education, and Wage Attainment," *American Journal of Sociology* 110, 123–60 doi:10.1086/386273

Jaeger, M., and A. Holm (2005) "Which Background Factors Matter More in Intergenerational Educational attainment: Social class, Cultural Capital or Cognitive ability? Random Effects Approach" *CAM (Centre for Applied Microeconometrics)*. University of Copenhagen

Jann, B. (2017) "Kernel Matching with Automatic Bandwidth Selection"

King, G., and R. Nielson (2019) "Why Propensity Scores Should Not Be Used for Matching," November, 1–20 doi:10.1017.pan.2019.11

Krueger, A., and O. Ashenfelter (1992) "Estimates of the Economic Return to Schooling from a New Sample of Twins," August 1. available at. Last updated December 17, 2020

Littnerová, S., J. Jarkovský, J. Pařenica, T. Pavlík, J. Špinar and L. Dušek (2013) "Why to Use a Propensity Score in Observational studies? Case Study Based on Data from the Czech

Clinical Database AHEAD 2006-09," *Cor et Vasa* 55, e383–90 doi:10.1016/j.crvasa.2013.04.001

Mine Çetinkaya-Rundel (2013) "Random Sampling vs. Random Assignment." YouTube Video *Duke University*

Mohnen, M. (2020) Matching *Applied Econometrics: University Of Ottawa*

"National Longitudinal Survey of Youth 1997 (NLSY97)," (2020). US Bureau of Labor Statistics available at. Last updated October 29, 2020

NORC (1987) "National Longitudinal Survey Questions" available at.

Oreopoulos, P. (2011) "Why Do Skilled Immigrants Struggle in the Labor Market? a Field Experiment with Thirteen Thousand Resumes," *American Economic Journal: Economic Policy* 3, 148–71 doi:10.1257/pol.3.4.148

"Pew Research Center," (2019) available at https://www.pewresearch.org/

Roksa, J. (2011) "Differentiation and Work: Inequality in Degree Attainment in U.S. Higher Education," *Higher Education* 61, 293–308

Spence, M. (1973) "Job Market Signaling," *The Quarterly Journal of Economics* 87, 355 doi:10.2307/1882010

"Stata Heckman Selection Model - Manuel" (2020) *Stata* Accessed December 20

UNDP (2020) "The next Frontier: Human Development and the Anthropocene Human Development Report 2020" *Human Development Reports*. United Nations Development Programme

Tucker, J., I. Thank, F. Aitsahlia, J. Demski, M. Donohoe, M. Flannery, B. Greene, M. Kirk, T. Omer, D. Schroeder and S. Tse (2011) "Selection Bias and Econometric Remedies in Accounting and Finance Research" *Forthcoming in Journal of Accounting Literature*

US Census Bureau (2019) "Industry and Occupation Code Lists & Crosswalks," November 14. available at. Last updated December 1, 2019

Yazbeck, M. (2019) Causality and Randomized Evaluations in Health Economics *University of Ottawa*

Appendix

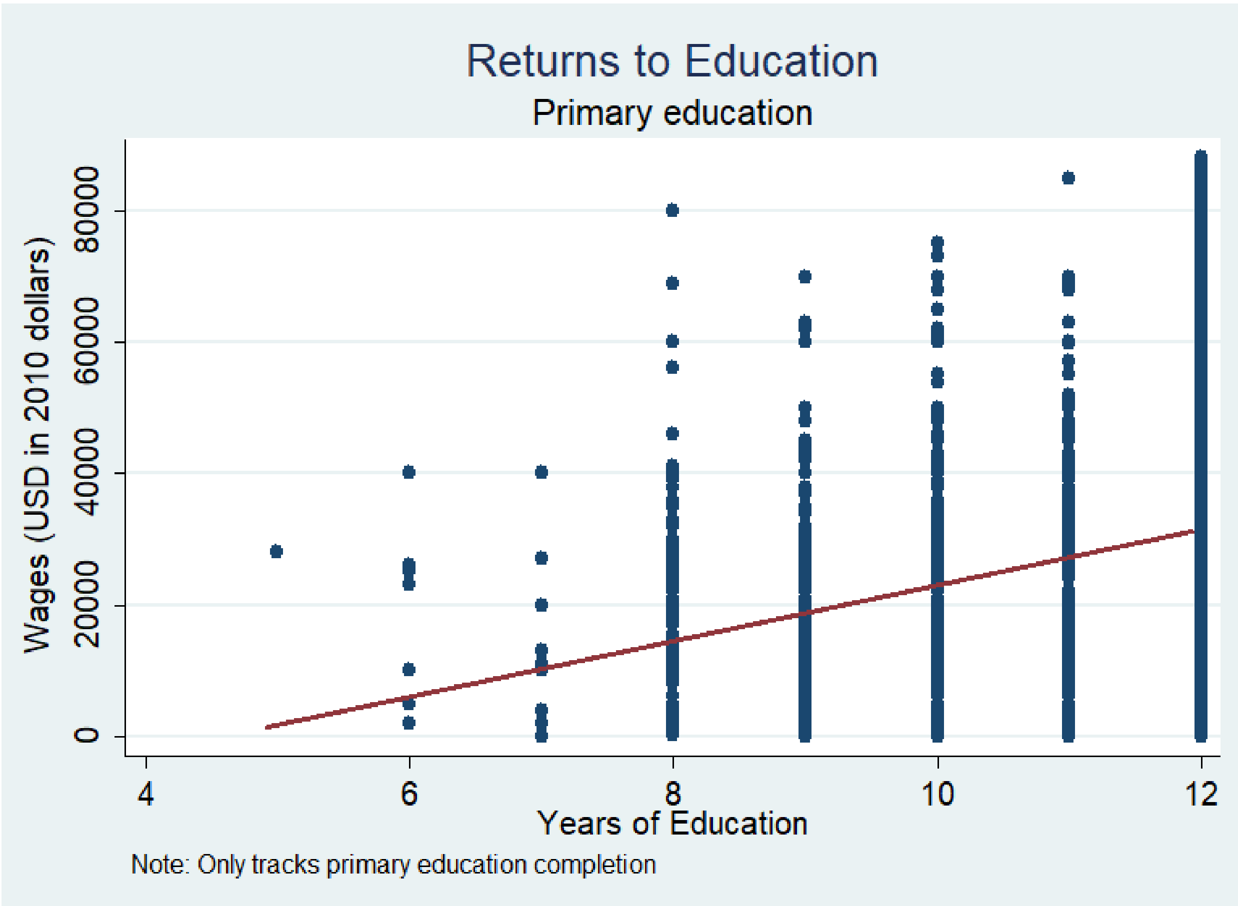


Figure 1 - Returns to Primary Education

Individuals who have not finished their primary education are earning less on average

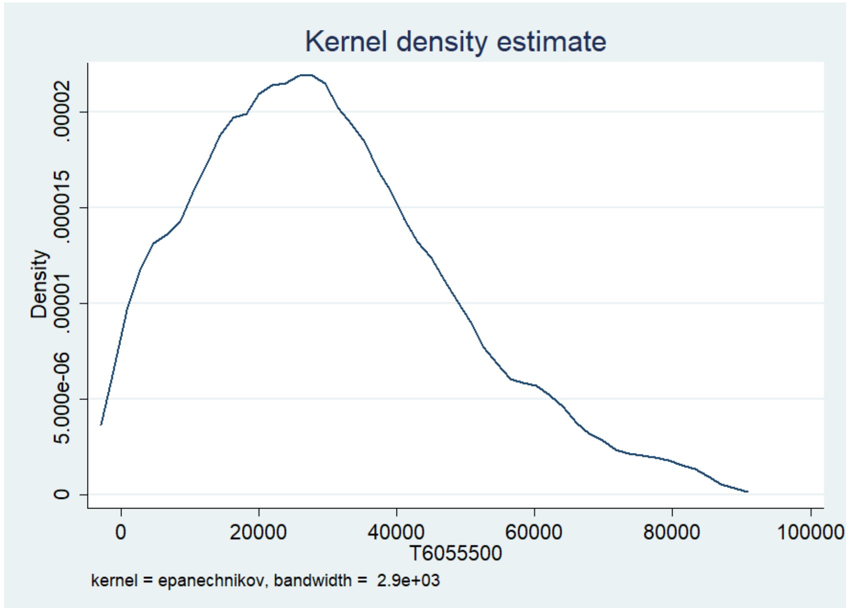


Figure 2 - Wages Distribution

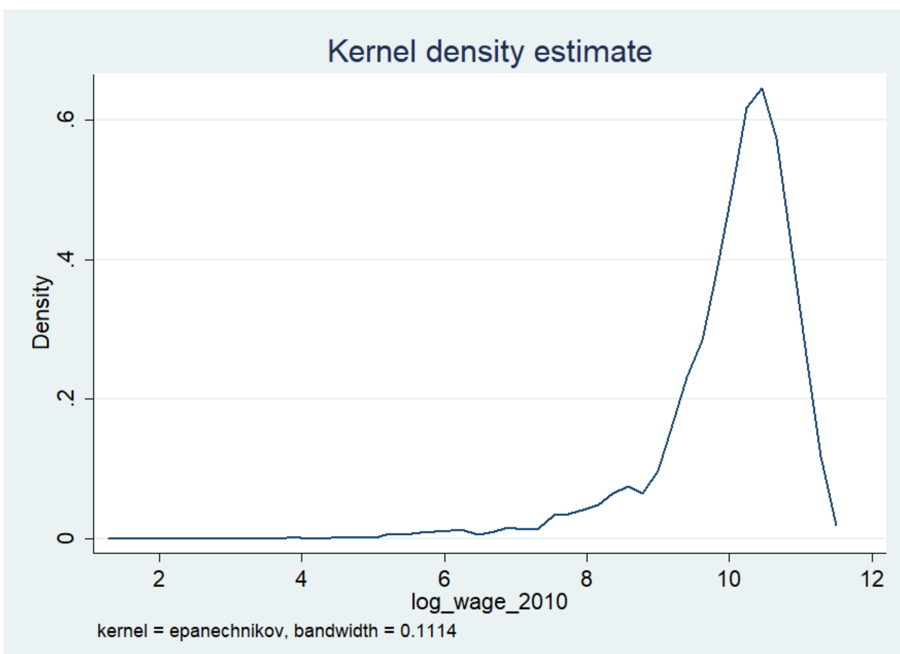


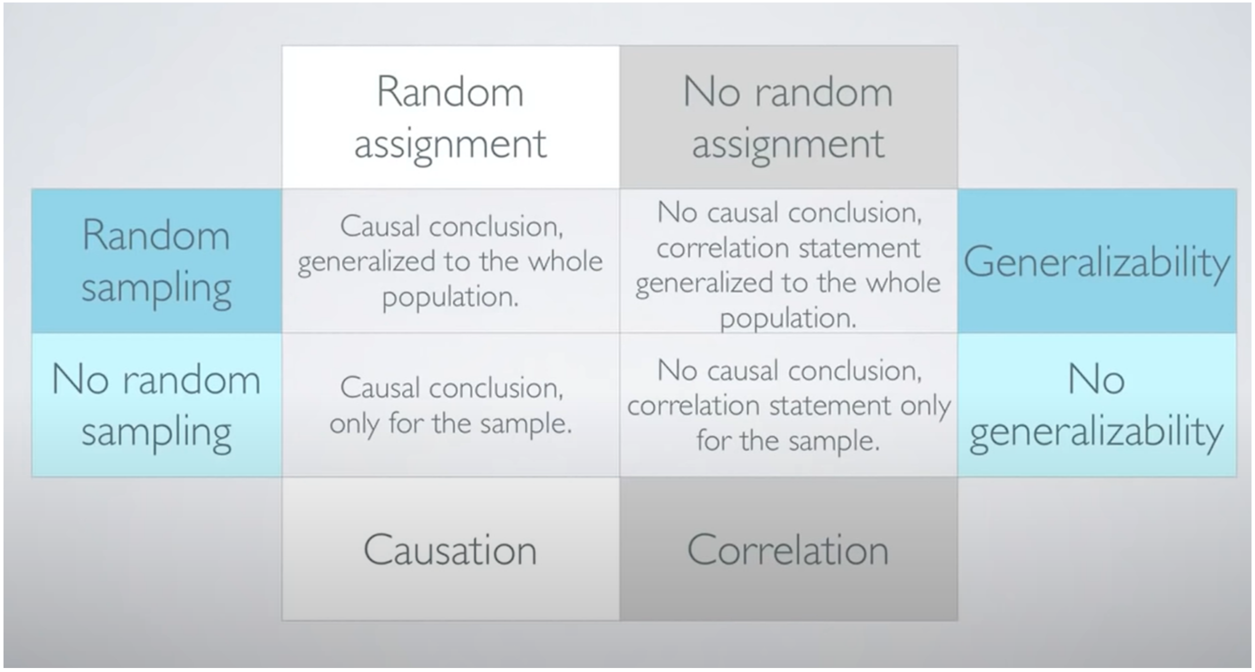
Figure 3 - Log Wages Distribution

. sum

Variable Legend: Continious Categorical Dummy

Variable	Obs	Mean	Std. Dev.	Min	Max	
ID_1997	8984	4504.302	2603.136	1	9022	Respondent ID number for consistency and privacy
SEX_1997	8984	1.48809	.499886	1	2	Female=1 male=2, much lower R_Sex fixes it to Female=1 Male=0
BDATE_M_1997	8984	6.555988	3.469429	1	12	Birth Month — Need to use both if I end up using it
BDATE_Y_1997	8984	1982.01	1.39687	1980	1984	Birth Year
Par_GROSS_~R	6588	46361.7	42143.5	-48100	246474	HH Gross Income from Parents
SAMPLE_~1997	8984	.7511131	.4323922	0	1	Sample or oversample
RACE_ET~1997	8984	2.787845	1.313645	1	4	Race /Ethinicity
AGE_2010	7479	27.89905	1.436392	25	31	Age
CENSUS_~2010	7422	2.706818	.9863668	1	4	Region
HIGHEST~2010	7440	2.308871	1.385705	0	7	Highest Degree
URBAN_R~2010	7415	.8095752	.4160135	0	2	Urban or Rural
SAMPLIN~2010	8984	215699.6	146234.5	0	1922275	Sampling Weights
College_GPA	720	410.1431	819.7229	0	9800	College GPA
Wages_2010	5147	29661.62	18380.69	0	88000	Wages for 2010
Height_feet	7307	5.183386	.430856	1	7	Height in feet — Need to use both if I end up using it
Height_inc~s	7295	5.298012	3.416442	0	12	Height in inches
DESCRIBE_W~T	7301	3.519929	.8168055	1	5	Describe Weight
GENERAL_HE~H	7473	2.307775	.9831366	1	5	General Health
Health_lim~k	7451	.0590525	.2357387	0	1	Did health inhibit work?
PS1_HARD_W~R	7443	1.878544	1.508384	1	7	Psychology: Hard worker? 1 = Strongly disagree 7 = Strongly Agree
PS2_Amount~k	7443	2.354964	1.733531	1	7	Amount of Work
PS3_Work_S~d	7389	6.26147	1.129347	1	7	Work Standart
PS4_Effort~k	7419	5.905648	1.216847	1	7	Effort at Work
PS5_Follow~s	7420	4.412399	1.92381	1	7	Following Rules
PS6_Breaki~s	7442	3.296291	2.18028	1	7	Breaking Rules
PS7_Suppor~d	7413	5.102253	1.590603	1	7	Support Rules and Traditions
PS8_Bendin~s	7411	4.60842	1.880134	1	7	Bending Rules
Occupation~y	6265	4454.828	2436.584	10	9990	Occupation Code
AGE_2013	7140	31.00098	1.437545	28	34	Age 2013
CENSUS_~2013	7072	2.710266	.980602	1	4	Census Region 2013
HIGHEST~2013	7114	2.417065	1.440295	0	7	Highest Degree 2013
URBAN_R~2013	7062	.8496177	.3790084	0	2	Urban Rural
wages_2013	5116	35243.38	22543.77	0	111131	Wages
weight_~2013	6925	185.5375	50.64451	1	474	Statistical Weight
GRIT_1_~2013	6924	3.376228	1.079327	1	5	Grit: New Ideas Distract From Old Ones 1=Strongly disagree 5=Strongly Agree
GRIT_2_~2013	6966	3.007608	1.215159	1	5	Setbacks Do Not Discourage Me
GRIT_3_~2013	6949	3.6765	1.100634	1	5	Short Term Obsessions
GRIT_4_~2013	7068	1.399689	.6806342	1	5	Hard Worker
GRIT_5_~2013	7015	3.616964	1.051007	1	5	Change Goals Frequently
GRIT_6_~2013	6988	3.748283	1.10506	1	5	Maintaining Focus
GRIT_7_~2013	7058	1.950553	.8923766	1	5	Ability To Finish Projects
General~2013	7130	2.319355	.9879218	1	5	General self reported health
health_~2013	7119	.0703751	.255796	0	1	Health inhibiting work?
college~2015	464	408.625	779.1125	0	9300	College GPA
Marital_St~s	7472	.5517934	.798033	0	4	Marital Status 2010
Educati~2010	7403	11.44144	1.145658	5	12	Education (Highschool) 2010
Educati~2013	7047	13.67107	2.986523	5	20	Education (Highschool) 2013
Union_member	5069	.1347406	.34148	0	1	Union member 2010
Job_satisf~n	5826	2.043083	1.027868	1	5	Job satisfaction 2010
TRAINING_P~m	1038	.2475915	.4318213	0	1	Training 2010
R_Sex	8984	.4880899	.499886	0	1	Gender
Urban_2010	7345	.7982301	.4013487	0	1	Urban or rural with third option of "I dont know" removed 2010
Urban_2013	7006	.8404225	.3662399	0	1	Urban or rural " 2013
parents_in~s	6588	2.094718	1.032411	1	5	Parents Income Status
Industry	6265	18.585	8.76201	1	33	Industry
gpa_2010	712	3.291615	.5885803	0	4	Gpa 2010 with fixed scale
gpa_2013	457	3.357593	.5467715	0	4	Gpa 2013

Figure 4 - Summary statistics



	Random assignment	No random assignment	
Random sampling	Causal conclusion, generalized to the whole population.	No causal conclusion, correlation statement generalized to the whole population.	Generalizability
No random sampling	Causal conclusion, only for the sample.	No causal conclusion, correlation statement only for the sample.	No generalizability
	Causation	Correlation	

Figure 5 Explanation of random assignment and sampling - Rundel M.C (2013)

```
. kmatch md R_Sex Union_member Industry_Category RACE_ETHNICITY_1997 GENERAL_HEALTH ( Wages_2010 ), att vce(boot) over( HIGHEST_DEGREE_2010 )
(HIGHEST_DEGREE_2010=0: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=1: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=2: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=3: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=4: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=5: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=6: computing bandwidth ... done)
(HIGHEST_DEGREE_2010=7: computing bandwidth ... done)

(running kmatch on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.....x..x...x..xxxxx.x.x....x..x.........x..x....    50

Multivariate-distance kernel matching          Number of obs      =      4111
                                               Replications       =        36
                                               Kernel             =      epan

Treatment   : R_Sex = 1
Metric      : mahalanobis
Covariates  : Union_member Industry_Category RACE_ETHNICITY_1997 GENERAL_HEALTH
Over groups : 0: HIGHEST_DEGREE_2010 = 0
              1: HIGHEST_DEGREE_2010 = 1
              2: HIGHEST_DEGREE_2010 = 2
              3: HIGHEST_DEGREE_2010 = 3
              4: HIGHEST_DEGREE_2010 = 4
              5: HIGHEST_DEGREE_2010 = 5
              6: HIGHEST_DEGREE_2010 = 6
              7: HIGHEST_DEGREE_2010 = 7
```

Matching statistics

	Matched Yes	No	Total	Controls Used	Unused	Total	Bandwidth
0 Treated	94	0	94	160	9	169	1.496497
1 Treated	161	0	161	213	11	224	1.276665
2 Treated	839	1	840	979	8	987	1.181718
3 Treated	182	0	182	138	2	140	1.669268
4 Treated	601	2	603	449	0	449	1.876547
5 Treated	138	1	139	84	0	84	1.989584
6 Treated	4	0	4	1	0	1	4.242641
7 Treated	19	1	20	12	2	14	3.224331

Treatment-effects estimation

Wages_2010	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
0 ATT	-9341.55	1527.187	-6.12	0.000	-12334.78	-6348.318
1 ATT	-2521.598	1463.737	-1.72	0.085	-5390.47	347.2739
2 ATT	-5924.133	770.9908	-7.68	0.000	-7435.247	-4413.019
3 ATT	-5535.253	2187.226	-2.53	0.011	-9822.136	-1248.369
4 ATT	-4375.571	940.3422	-4.65	0.000	-6218.608	-2532.534
5 ATT	-2504.82	2493.351	-1.00	0.315	-7391.699	2382.059
6 ATT	20092	10842.52	1.85	0.064	-1158.956	41342.96
7 ATT	12726.85	8006.346	1.59	0.112	-2965.305	28418.99

Figure 6 - Kernel Matching - Average treatment effects by Highest Earned Degree Subcategory

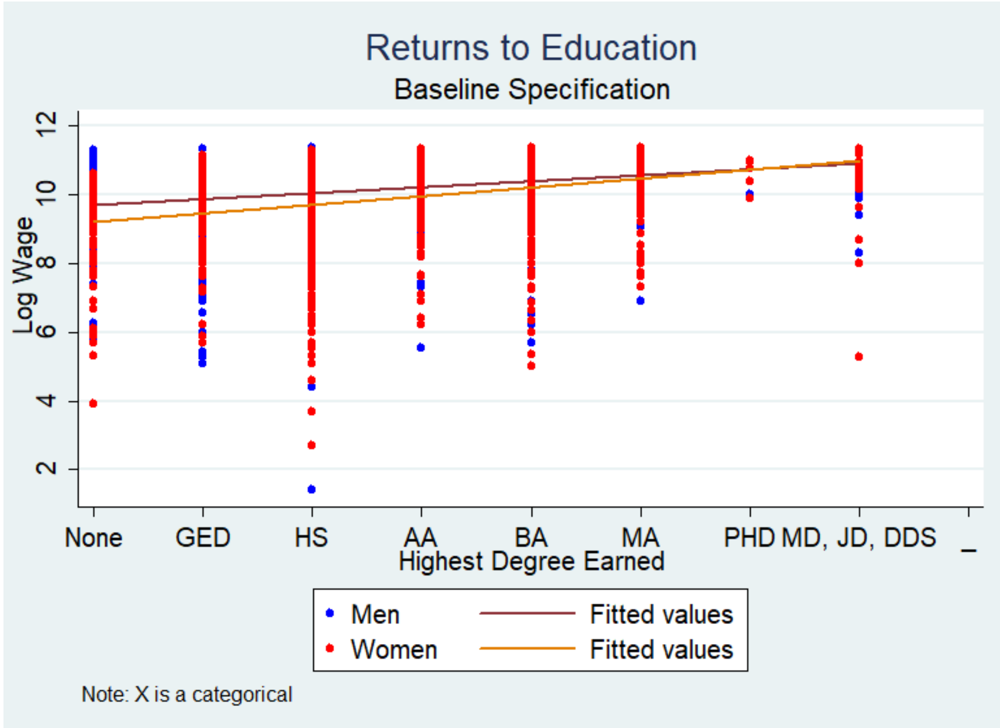


Figure 7 - Returns to Education categorised by gender

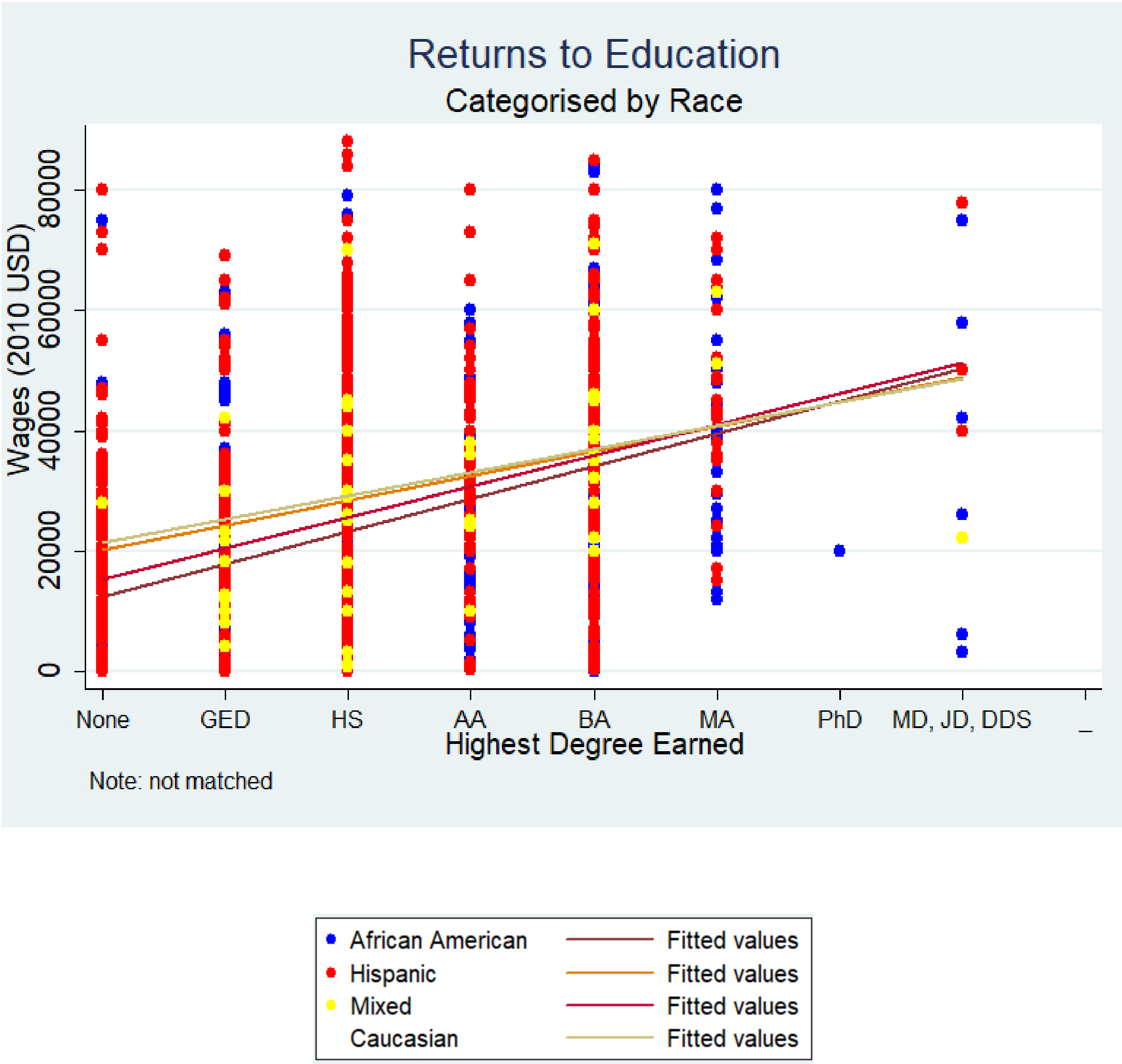


Figure 8 - Returns to Education categorised by race

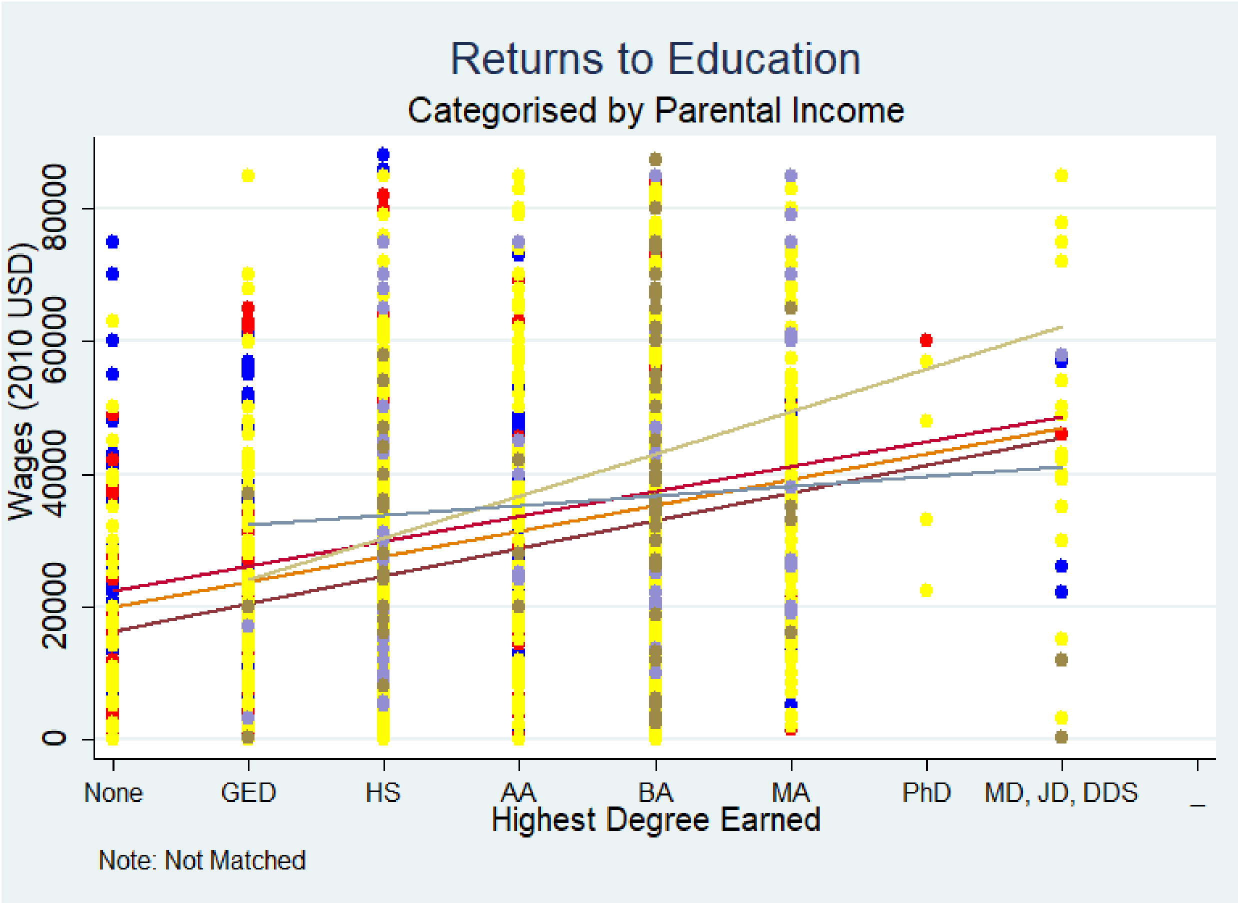


Figure 9 - Returns to Education categorised by parental income