Contents

MEASURING THE RETURNS TO SCHOOLING ACROSS TIME -THE PROFESSIONAL ECONOMIST

A Short Paper
Presented to Dr. Brian Duncan,
Department of Economics
University of Colorado
Denver, CO. 80204, USA

In Partial Fulfillment
of the Requirements for Econometrics 1

By
Richard A. Boateng

December 2020

1. INTRODUCTION

What is the return to education and is it changing over time? These are questions that keep coming up whenever a country, state, investors in education and family heads try to make decisions on 'educational investments'-a key part of growth. These decisions are continually being made at the state and national levels involving policies, for example to increase college enrollment rates as well as by institutions such as the World Bank (1995) to make education sector loans. These decisions normally apply to the education system, and what is most relevant is the expected rate of return to society on investment in education as a whole that does take institutional costs into account (Arias & McMahon, 2001). Majority of modern literature on this the topic suggest that 'better' education is highly positively correlated with higher wages, more employment and prestigious jobs. Is this relationship changing over time? Although most literature supports the changing returns to education over time, one must practice caution drawing strong causal conclusions about the return to education for some specific occupations and how that changes over the course of time. This paper analyzes the return to education for economist. The continues

development of new technology and availability of more data is expected to increase the effectiveness, efficiency and impact of the work of an economist, and thus the rise in returns to education for economist over time.

The Data set was extracted from The Integrated Public Use Microdata Series (IPUMS-USA) website from the American Community Survey (ACS) for the years 2010 to 2018. The occupation here is economist - a professional working in one of the many fields of economics or having an academic degree in economics. The sample is limited to only the employed. The average wage in the sample is 57.17 dollars, 61.11 and 49.42 dollars for men and women respectively. The age was trimmed to include only individuals between the ages 25 and 59 years.

The data consist of rich information about individuals, states, years and a wide range of corresponding characteristics. The data initially included elementary, middle, secondary and graduate economist. We eliminated from our data set anyone who works 26 weeks or less out of the year and anyone who works less than 20 hours per week or more than 60 hours per week leaving us 1,860 number of observations with 1232 males and 628 females. Using Instrumental Variable (IV), Two Stage Least Square (2SLS) techniques, and fixed effect modeling with interactions, we try to estimate the causal effect and dynamics of level of education on log wages of economist using quarter of birth as our instrument for level of education. We found that controlling for age, state and year fixed effects, an additional year of schooling for economist is associated with 7.8% increase in hourly wages. We estimated the highest year of return to education being a 11.6% returns to hourly wages for an additional

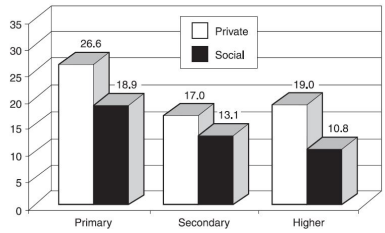
year of schooling in the year 2015 and 11.1 % in 2017. The returns to education for the remaining years hovers between 8% and 9%. Thus, we did not find any trend in the returns to education for economist across time although the return differ across time.

2. BACKGROUND

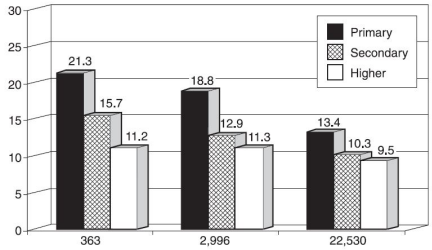
2.1 Literature

There has been numerous studies that attept to measure the causal effect of education on earnings. These studies varies in objectives, scope, data types and even methodology. For example (Arias & McMahon, 2001) describes their objective as to calculate pure internal rates of return that are relevant to economy-wide investment decisions. Their study focused on the returns to education over time for given "synthetic cohorts" of workers, including recent graduates, where each cohort is comprised of individuals who graduate in the same year, are of the same gender, and share the same educational attainment (Arias & McMahon, 2001). Here we use annual CPS data for males and for females separately, as well as institutional costs annually, to recompute a dynamic or expected dynamic rate of return each year from 1967 through 1995.They employed the OLS and GLS methods in achieving this aim (Arias & McMahon, 2001). Another focused type of studies that runs across the literature are those that use Mincer earnings function and cross sectional data to focus on racial

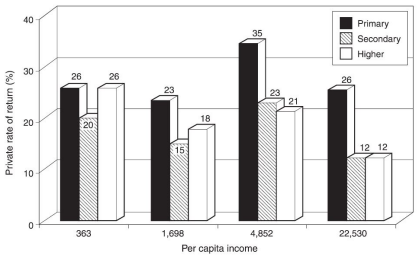
differences or on corrections for endogeneity of schooling and measurement error in education (Arias & McMahon, 2001). Uusitalo (1999) estimates the return to education in Finland using an individuallevel data set that also includes ability measures and information on family background. They concluded that ability test scores have a strong effect on the choice of education and on subsequent earnings. Estimating the return to education with no information on ability leads to an upward bias in the estimates. However, this bias is more than offset by a downward bias caused by endogeneity or measurement error. They employed the instrumental variables estimates technique that utilize family background variables as instruments to produce estimates of the return to schooling that are approximately 60% higher than the least squares counterpart (Uusitalo, 1999). Patrinos & Psacharopoulos (2002) found that overall, the average rate of return to another year of schooling is 10%. They also found the returns to education on country level income. Their findings suggested the highest returns are recorded for low-income and middle-income countries. Latin America and the Caribbean have the highest average returns to schooling followed by the sub-Saharan Africa region. Returns to schooling. However, they found that the returns are lower in the high-income countries of the OECD. He also found that average returns to schooling are lowest for the non- OECD European, Middle East and North African group of countries. Figure 2.1.1 below summarizes his findings that suggest the returns to education may change overtime.



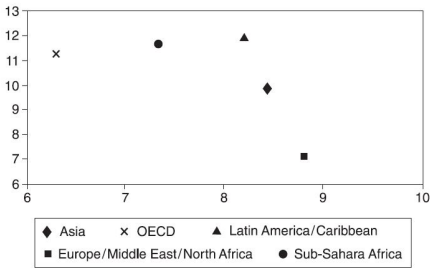
(a) Returns to investment in education by level, latest year.



(b) Social returns to investment in education by income level.



(c) Private returns to investment in education by income.



(d) Mincerian returns and mean years of schooling.

Figure 2.1.1: The above graphs (a) Returns to investment in education by level, latest year, (b) Social returns to investment in education by income level, (c) Private returns to investment in education by income and (d) Mincerian returns and mean years of schooling are from the works of Patrinos & Psacharopoulos (2002).

2.2 Data

The Data set was extracted from The Integrated Public Use Microdata Series (IPUMS-USA) website from the American Community Survey (ACS) for the years 2010 to 2018. The sample is a repeated cross-sectional data. The The data initially included Elementary, Middle, secondary and graduate economist. We eliminated from our data set anyone who works 26 weeks or less out of the year and anyone who works less than 20 hours per week or more than 60 hours per week leaving us 1,881

number of observations with 1247 males and 634 females. The unit of observation is individual in a state at a particular year. We found overall average wage is 57.17 and the average age is 41.5 years. Mean wage for men is 61.11 dollars per hour and Mean wage for women is 49.42 dollars per hour. 'exp' is a person's potential labor market experience. The variable "exp" equals the person's age minus their years of schooling minus six. The data is a repeated cross sectional data with the year the data was collected and the state being the unit of observation. The sample was limited to individuals who work 26 weeks or more, and 20 to 60 hours per week.

Table 8.0.1 shows all the variables included in the study. For example the variable "years of schooling" measures the level of education attained. It is an ordinal variable with the lowest and highest level of educational attainment being high school and PhD. respectively. The variable wage measures hourly wages of individuals in the data calculated from the number of weeks worked equal to a person's yearly wage or salary income divided by the total number of hours a person works in a year. These are calculated from income per week, number of hours worked per week, and the number of weeks worked which is included in the ACS data. The female variable is an indicator variable which is equal 1 when an individual is a female and zero for males. The Statefip variable is a two digit code that uniquely identifies each of the states in the US. The 'racecat5' variable is a categorical variable representing each of the five races : 'Whites', 'Blacks', 'Hispanics', 'Asians' and 'Natives/Other races'. The 'QOB' variables are dummy variables indicating the quarter of year in which an individual was born.

3. Basic Patterns in the Data

Figure 8.0.1 Panel A is line graph showing the average years of schooling by year. We observe that the lowest average number of years spent in school occurs in 2004 whiles the highest average number years spent in school occurs in 2018 which is about 17.95 years and a 18.16 years respectively. There is no obvious trend in the graph. Figure 8.0.1 Panel B is a line graph showing the average wage received by economist by year. The highest average wage received occurs in 2017 while the lowest wage received occurs in 2010 with wages about 51 dollars and 63 dollars respectively. Average wages received by economist appears to be an increasing over time.

4. Empirical Specification

The first baseline OLS regression used in this analysis is:

$$\ln(wage_{ist}) = \alpha + \pi schoolyr_{ist} + X_{ist}\beta + \mu_s + \tau_t + \epsilon_{ist}, \tag{4.0.1}$$

$$\ln(wage_{ist}) = \alpha + \phi_t schoolyr_{ist} + X_{ist}\beta + \mu_s + \tau_t + \epsilon_{ist} \tag{4.0.2}$$

where $\ln(wage_{ist})$ is the wage of individual i living in state s in year t. The vector X_{ist} includes controls for the five race/ethnicity categories, a female indicator variable, and integer age fixed effects. Some regressions will also include state fixed effects, μ_s, and year fixed effects,τ_s. All regressions will estimate heteroskedastic robust standard errors clustered at the state level. To estimate Equation 4.0.1, we use both OLS technique and IV regression technique. Note that the coefficient (π) from Equation 4.0.1 is a log-linear model so it will be interpreted as the average wage increase by $(\pi \times 100)\%$ for an additional year of schooling. Adding age fixed effect to the model controls for the fact that years of schooling typically increases with age thus omitting

age fixed effect will lead a positive bias on the coefficient of schooling. Moreover, adding the state fixed effect controls for all the observable and un-observable unique state characteristics that does not vary across time. There may be some characteristics in one state that makes individuals from that state earn much more or less than other states. These fixed state characteristics is controlled for by the state FE.

Estimating Equation 4.0.1 using OLS will lead to bias of the π in Equation 4.0.1. The OLS estimates will ignore the fact that the data in question is a repeated cross sectional organized by time and units. Thus it pools the observation and conduct the regression in a naive way with the assumption that intercepts are the same for each unit and for each time period, the coefficient for each units are the same and constant over time that the structure of the error is the same for each unit and for each time period. Specifically, adding age fixed effect to the model controls for the fact that years of schooling typically increases with age thus omitting age fixed effect will lead a positive bias on the coefficient of schooling. Additonally, adding the state fixed effect controls for all the observable and un-observable unique state characteristics that does not vary across time. There may be some characteristics like minimum wage in one state that makes individuals from that state earn much more or less than other states and does not vary across time. These fixed state characteristics is controlled for by the state FE.

Using the quater of birth variable as an instrument for years of schooling remains a relevant instrument because it determines when a person starts kindergarten, which in turn determines what grade a person is in when he or she can legally drop out of

school which in turn can be argued that students who are not allowed to drop out of school before they start their senior year are more likely not to drop out leading to increase in the number of years spent in schooling. Moreover, quater of birth can be seen as a variable that is randomly assigned and and it assignment had nothing to do with the level of education attained. Quater of birth is a valid instrument since it does not directly affects wages but only through the level of education attained variable. One drawback to using quarter of birth as an instrument may a situation when some parents deliberately time their children's birth to occur at a particular time of year. Doing so may not make the quarter of year variable randomly assigned. Another potential problem with quarter of birth as an instrument is that it may cause changes in other variables, which in turn may cause changes in wages making it invalid. But we assume these situation were absent or in the worse case minimal which could not affect its random assignment.

The coefficient on Equation 4.0.2 (ϕ_t), estimates separate returns to schooling for each year by estimating different coefficients for the level of education attained per each year. Thus, ϕ_t can be interpreted as the $(\phi_t \times 100)$ percent increase in wages for an additional year of schooling in year t. For economist, the continues development of new technology and availability of more data is expected to increase the effectiveness, efficiency and impact of their work, and thus the rise in the return to education for economist over time. Thus I expect ϕ_t to rise over time. Note that Equation 4.0.2 does not estimate different coefficients for age, race/ethnicity, female, and the state fixed effects by year as separate regressions by year would.

5. Regression Results

The results in Table 8.0.2 shows the first stage regression in the 2SLS regression in Equation 4.0.1. Heteroskedasticity-robust standard errors clustered at the state level are shown in parentheses. The instruments are the quarter of birth instrument with 4th quarter serving as the reference category. The race/ethnicity variables are mutually exclusive and exhaustive with white serving as the reference category. The dependent variable is the level of education attained. Column 1 is without age, state and year fixed effect. The coefficient of the first quarter is 0.127, which is interpreted as being born in the 1st quarter increases your level of education attained by 0.127 more than a person born in the fourth quarter. Similarly, a person born in the 3rd quarter has 0.161 more education than a person born in the fourth quater.

Column 2 on the other hand has age, state and year fixed effect. The coefficient on the first quarter is 0.18, which can be interpreted as controlling for age, state and year characteristics, a person born in the first quarter gets 0.18 more education than a person born in the fourth quarter. Similarly, a person born in the 3rd quarter gets 0.183 more education than a person born in the 4th quarter controlling for age

state and year. The coefficient on the female dummy variable is -0.142, which can be interpreted as females gets 0.142 less education than men with the same age, state and year characteristics. Note that the F statistics in both regressions are very small with non of the coefficients statistically significant. This implies the the instruments used here (quarter of birth) does not work as a good instrument, likely because the sample was limited in size and scope (only economist).

Table 4.0.2 presents the results of the OLS and the IV estimates of the returns to education presented in Equation 4.0.1. The dependent variable is the log of hourly wage. Column (1) presents the OLS results without state and year fixed effect. The coefficient on the level of education attained is 0.101 which can be interpreted as controlling for age, a one more year of schooling attained is associated with 10 percent increase in hourly wages which is statistically significant at the 5 percent confidence level. Column (2) presents the OLS results with only state and age fixed effect. The coeficient on level of education attained is 0.078 which can be interpreted as controlling for state and age fixed effect, an additional year of schooling is associated with 7.8 percent increase in wage. Column (3) controls for age, state and year fixed effect. Adding the state and year fixe effect decreased the coefficient on years of education attained from 0.101 to 0.078(comparing columns (1) and (3)) representing a 2.3 percentage points decrease in the returns to education. Adding the state, age and year controls to column(5), the IV regression, decreased the coefficient on the level of education attained from 0.296 to 0.274 which is about a 2.2 percentage points decrease. Note that the instrument used does not appear to be relevant. Thus, there

is a significant difference between the level of education attained coefficient estimates in column (3) and column (5). The coefficient in for level of education variable in column (5) is expected to be smaller than that in column(3) but it did not.

Table 8.0.4 presents the OLS estimates of the returns to schooling by year from 2010 to 2018. The coefficients ϕ_t are the results Equation **??** representing the interactions between the year and the level of education attained variable. All the coefficients are statistically significant at the 0.05 singnificant level with the exception of the returns to 2011 [1]. The returns to education hovers around 8.5 percent increase in wages across all years with a spike in 2015 and 2017 to about 11 percent increase in wages. Using a joint significant F-test, yielding a F- statistic of 2.11, the null hypothesis was rejected at 0.05 significant level. Thus, our model performs well. Thus the return to education in 2011 is 8.4 percent and 2015 is 11.6 percent which represents an increase. Thus, these coefficients are meaningfully different.

[1] An alternate way to estimate the returns to schooling by year is to estimate regression 8.0.4 using separate regression for each year. This was estimated but not presented in this research, and the results is identical to the one shown in the paper.

6. Robustness Checks

Table 8.0.5 presents the results for robustness check. The first column(1) shows the OLS estimation with the outcome variable as levels of wage regressed on all the independent variables described in Equation 4.0.2. Column 2 of Table 8.0.5 is an OLS estimates of Equation 4.0.2 with the dependent variable as log of wage regressed on a standardized level of schooling and year interaction variable. The results in column 3 is similar to column 2 except that the level of schooling and the year interaction variable is centered.

7. Conclusion

Data driven studies across a wide range of occupations have confirms this. Consequently, the returns to education for economist is expected to change over time. We first use a panel data analysis technique by making use of pooled regression. We then used an instrumental variable (IV) and two stage least square (2SLS) technique, using quater of birth as an instrument, to extract and isolate the exogenous variations in the level of education attained that was caused by random assignment of the quater of birth variable. We proceed to use Fixed Effect (FE) regression techniques, controlling for age, year, state, race and gender in different mixtures, to determine the causal effect of level of education attained on the wages. In order to establish the dynamics of the effect of level of education attained on the wages of economist, we made use of interactions between the level of education attained and the year variables. We found that controlling for age, state and year fixed effect, an additional year of schooling increases hourly wages by 8.11 percent for an economist. Although we did not find any specific trend, returns to education is not constant over time for the economist profession. An Ideal research will be to use longitudinal data with level of education

randomly assigned or using longitudinal life histories of earnings data for individuals, but longitudinal data sets are rare, especially over long periods of time that extend into current years. Thus, with a repeated cross sectional data the econometrics tools employed in this analysis, I am much confident about the results.

8. Appendix

Figure 8.0.1: Average years of schooling, by year.

Variable	Description
year	Cencus year birth.
statefip	State (FIPS code).
sex	Sex.
age	Age.
birth_qtr	Quarter of birth.
hispanic	Dummy variable equal to one for Hispanics, zero for non Hispanic.
black	Dummy variable equal to one for Blacks, zero for non-Blacks.
asian	Dummy variable equal to one for Asians, zero for non Asians.
native	Dummy vairiable equal to one for Natives, 0 otherwise.
white	Dummy variable equal to one for Whites, zero for non Whites.
othrace	Dummy variable equal to one all other categories not in the previous categories.
race_cat5	Race/Ethnicity Category.
female	Female indicator variable.
wage	The workers wage.
lnwage	Log of wage.
years of schooling	Level of education attained.
QOB1	First Quarter.
QOB2	Second Quarter.
QOB3	Third Quarter.
QOB4	Fourth Quarter.

Table 8.0.1: Table 1: Variable Description

	(1) Level of ecucation attained	(2) Level of ecucation attained
(Instruments)		
First Quarter	0.127	0.180
	(0.109)	(0.106)
Second Quarter	0.133	0.146
	(0.108)	(0.108)
Third Quarter	0.161	0.183
	(0.109)	(0.107)
Female indicator variable		-0.142
		(0.075)
(Race/Ethnicity)		
Black		-0.308
		(0.176)
Asian		0.693***
		(0.101)
Hispanic		-0.043
		(0.149)
Native/Other_race		-0.214
		(0.264)
Constant	17.985***	16.380***
	(0.083)	(0.178)
State fixed effects	No	Yes
Year fixed effects	No	Yes
Age fixed effect	No	Yes
Mean of dependent variable	18.1	18.1
R-squared	.0015	.0933
Sample size	1860	1860
F-stat on instruments	.815	1.23

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: 2010 - 2018 American Community Survey (ACS).

Notes: The sample includes working men and women between the ages of 25 and 59. The dependent variable is level of education attained. Heteroskedasticity-robust standard errors clustered at the state level are shown in parentheses. Fourth quarter of birth is the reference category. The race/ethnicity variables are mutually exclusive and exhaustive with white serving as the reference category.

Table 8.0.2: Table 2: Years of schooling first stage IV regressions

	(1-OLS) log of wage	(2-OLS) log of wage	(3-OLS) log of wage	(4-IV) log of wage	(5-IV) log of wage
Level of ecucation attained	0.101***	0.079***	0.078***	0.296	0.274
	(0.010)	(0.010)	(0.010)	(0.237)	(0.185)
Female indicator variable	-0.155***	-0.136***	-0.130***		-0.130**
	(0.026)	(0.025)	(0.025)		(0.040)
(Race/Ethnicity)					
Black	-0.249**	-0.272***	-0.274***		-0.196*
	(0.080)	(0.082)	(0.080)		(0.086)
Asian	0.014	-0.018	-0.022		-0.104
	(0.042)	(0.040)	(0.040)		(0.134)
Hispanic	0.002	-0.022	-0.036		0.009
	(0.044)	(0.043)	(0.043)		(0.048)
Native/Other_race	-0.189*	-0.171*	-0.163*		-0.156
	(0.087)	(0.079)	(0.080)		(0.103)
Constant	1.606***	1.580***	1.497***	-1.487	-1.241
	(0.195)	(0.213)	(0.212)	(4.295)	(3.062)
State fixed effects	No	Yes	Yes	No	Yes
Year fixed effects	No	No	Yes	No	Yes
Age fixed effect	Yes	Yes	Yes	No	Yes
Mean of dependent variable	3.86	3.86	3.86	3.86	3.86
R-squared	.267	.342	.353	.	.0853
Sample size	1860	1860	1860	1860	1860

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: 2010 - 2018 American Community Survey (ACS).

Notes: This table reports the OLS and IV estimates of Equation 4.0.1. The sample includes working men and women between the ages of 25 and 59. The dependent variable is the log of hourly wage. Heteroskedasticity-robust standard errors clustered at the state level are shown in parentheses. The race/ethnicity variables are mutually exclusive and exhaustive with white serving as the reference category. See Table 8.0.2 for the first stage IV estimates.

Table 8.0.3: Table 3: OLS and IV estimates of the returns to schooling

	(1) log of wage
Year 2010 X schoolyr	0.084***
	(0.012)
Year 2011 X schoolyr	0.033
	(0.021)
Year 2012 X schoolyr	0.084**
	(0.026)
Year 2013 X schoolyr	0.072**
	(0.025)
Year 2014 X schoolyr	0.083**
	(0.027)
Year 2015 X schoolyr	0.116***
	(0.022)
Year 2016 X schoolyr	0.074**
	(0.025)
Year 2017 X schoolyr	0.111***
	(0.026)
Year 2018 X schoolyr	0.085***
	(0.017)
State fixed effects	No
Year fixed effects	Yes
Age fixed effect	Yes
Race fixed effect	Yes
Female Indcator	Yes
Mean of dependent variable	3.86
R-squared	.357
Sample size	1860

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: 2010 - 2018 American Community Survey (ACS).

Notes: This table reports the OLS estimates of Equation 4.0.2. The sample includes working men and women between the ages of 25 and 59. The dependent variable is the log of hourly wage. Heteroskedasticity-robust standard errors clustered at the state level are shown in parentheses.

Table 8.0.4: Table 4: OLS estimates of the returns to schooling, by year

	(1) wage	(2-STD) log of wage	(3-Centered) log of wage
Year 2010 X schoolyr	4.269***	0.486***	0.084***
	(0.835)	(0.071)	(0.012)
Year 2011 X schoolyr	2.124	0.195	0.033
	(1.622)	(0.122)	(0.021)
Year 2012 X schoolyr	5.302**	0.476**	0.084**
	(1.904)	(0.148)	(0.026)
Year 2013 X schoolyr	4.090**	0.419**	0.072**
	(1.492)	(0.146)	(0.025)
Year 2014 X schoolyr	4.684*	0.443**	0.083**
	(1.754)	(0.146)	(0.027)
Year 2015 X schoolyr	6.545***	0.642***	0.116***
	(1.732)	(0.122)	(0.022)
Year 2016 X schoolyr	4.169**	0.439**	0.074**
	(1.490)	(0.145)	(0.025)
Year 2017 X schoolyr	8.122***	0.650***	0.111***
	(1.898)	(0.152)	(0.026)
Year 2018 X schoolyr	6.314***	0.480***	0.085***
	(1.101)	(0.098)	(0.017)
State fixed effects	No	No	No
Year fixed effects	Yes	Yes	Yes
Age fixed effect	Yes	Yes	Yes
Race fixed effect	Yes	Yes	Yes
Female Indcator	Yes	Yes	Yes
Mean of dependent variable	57.2	3.86	3.86
R-squared	.258	.357	.357
Sample size	1860	1860	1860

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: 2010 - 2018 American Community Survey (ACS).

Notes: This table reports the OLS and IV estimates of Eq. (1). The sample includes working men and women between the ages of 25 and 59. The dependent variable is the log of hourly wage. Heteroskedasticity-robust standard errors clustered at the state level are shown in parentheses. The race/ethnicity variables are mutually exclusive and exhaustive with white serving as the reference category. See Table 3 for the first stage IV estimates.

Table 8.0.5: Table 5 – Robustness Checks

References

Arias, O., & McMahon, W. W. 2001. Dynamic rates of return to education in the us. *Economics of Education Review*, *20*(2), 121–138.

Patrinos, A. H., & Psacharopoulos, G. 2002. *Returns to investment in education: a further update*. The World Bank.

Uusitalo, R. 1999. Return to education in finland. *Labour Economics*, *6*(4), 569–580.