Earnings Returns to STEM Graduate Degrees

Differences between undergraduate populations

Carnagua, Stephen Andrew

5/2/2014

For fulfillment of the requirements for the Spring 2014 **E582** *course,* **Applied Microeconomics II**

Introduction

It is well established that higher education is associated with an increase in lifetime earnings (see for example, U.S. Census Bureau, 2013). An additional body of work concerns earnings premiums associated with specific undergraduate majors (Arcidiacono, 2004) or level of schooling (bachelor's, master's, Ph.D., professional school) (Craft, 2003). However, there is little or no research on the value of graduate degrees conditional on the undergraduate degree. Specifically, I hypothesize that a graduate degree in a Science, Technology, Engineering, or Mathematics (STEM) field is more profitable, in terms of lifetime earnings, to an individual who received a bachelor's degree in a non-STEM field.

In order to detect and measure this difference, I analyze earnings data between from the 2010 National Survey of College Graduates, conducted by the Census Bureau for the National Center for Science and Engineering Statistics. I examine differences in earnings outcomes between groups of individuals who have graduate degrees in STEM fields, distinguishing between undergraduate major classifications. In this paper, I briefly summarize previous related research, describe the available data, fit a multiple linear regression model, and analyze the results.

Literature Review

The focus of my literature review is to learn what techniques others have used to answer questions similar to mine. To that effect, I pay special attention to control and outcome variables selected by researchers and model choices. Fortunately, previous studies suggest a consistent list of control and outcome variables for my analysis.

Variable selection

In his survey of work on the effect of education on earnings, Card (1999) observed that a fundamental aspect of research on the topic is that "the return to education is not a single parameter in the population, but rather a random variable that may vary with other characteristics of individuals" (p. 1803). This heterogeneity helps explain why it is hard to pin down an exact measure of the return to education, and it serves as a useful starting point when addressing a new research question. Many researchers have studied how different educational outcomes affect earnings. Studies typically employ regression models, of the form:

$$log(earnings) = \alpha + \beta X + \gamma S + \varepsilon ,$$

where X is a vector of control variables, and S is the variable or variables used to capture schooling, such as years of schooling, type of degree obtained, prestige of institution, etc.

Most studies use the natural logarithm of earnings (hourly, weekly, or annual) as the dependent variable. This is by far the most popular choice, and Heckman and Polachek (1974) demonstrated that the log of earnings is the most appropriate form, using the Box-Cox transformation. Card (1999) notes several advantages of using log wages, including the fact that "the distribution of log earnings (especially log hourly wages) is surprisingly close to a normal distribution" and "perhaps as important as any consideration, the log transformation is convenient for interpretation" (p. 1808).

Common control variables include demographics, such as age, race, gender, or marital status (see Arcidiacono, 2004; Craft 2003; Brewer, 1999; and Link, 1975, to name a few); measures or proxies for ability, such as high school GPA (e.g., Brewer, 1999), tests on standardized tests such as the ACT or SAT (e.g., Arcidiacono, 2004), or whether an individual has published an article or book (e.g., Link, 1975); and family background variables, such as parents' education (e.g., Brewer, 1999). Researchers often try to find proxy variables for innate ability (e.g., Arcidiacono), because, as Hunt (1963) pointed out, it is difficult to distinguish between factors which lead to further or "better" education and the returns of further education or more prestigious education.

Techniques/models used

Most research on earnings and education relies on ordinary least squares regression. In fact, every study I have come across has used OLS, at least in part. Craft & Baker (2003) compared OLS and Tobit models for earnings of lawyers by degree, because earnings were top coded at $150,000 in their data. Interestingly, the Tobit results were virtually identical to those of OLS, meaning the top coding did not affect the results much, partially because there were few cases where earnings were over $150,000 and partially because those cases that were top coded did not substantially deviate from the OLS predicted values. They did find heteroskedasticity and simply corrected for it by reporting robust standard errors.

Sometimes instrumental variables are employed to account for simultaneous determination of measures of ability and schooling choice. Parents' education, for example, has often been chosen because it explains 30% of the variation in education in American adults (Card, 1999). Card notes that instrumental variables provide higher estimates of the return to schooling. However, he examines biases theoretically and empirically with National Longitudinal Survey data and concludes that appropriate instrumental variables have yet to be identified, and specifically that family background is inadequate, because the instrumental variables estimator is almost certainly biased.

Researchers have also suggested using twin studies, sibling comparisons, or father-son/mother-daughter pairs (Card, 1999). By directly comparing family members, researchers can theoretically control for many unobserved personal background variables and account for individual heterogeneity in returns to education. Card suggests these studies contain less ability bias than standard OLS models. He determines that these models have their own sets of biases, however, and OLS is probably adequate.

Every article examined used survey microdata to estimate the effect of education on individuals' earnings. For example, Craft(2003) used a cross section survey of lawyers, and Brewer et al. (1999) and Arcidiacono (2004) both used National Longitudinal Survey (NLS) data. Brewer et al. utilized longitudinal data, unlike most other researchers, and were therefore theoretically able to control for labor market cohort effects. In one section of his review, Card (1999) lists a variety of articles which controlled for family background, using NLS data, the General Social Survey of adult household heads, a survey of Finnish male veterans, and three twin studies.

Findings and estimation issues

While the question of differences in earnings premium to a graduate degree by undergraduate degree type has apparently not been taken up yet, previous work on earnings returns to various education outcomes is instructive. Here I briefly summarize a few works on earnings differences, their findings, and mention some of the issues which are germane to my research question.

First, Arcidiacono (2004) studied earnings returns to different undergraduate majors. He attempted to control for endogeneity issues by accounting for ability sorting in a separate model; because higher ability students may gravitate toward certain majors, it is difficult to differentiate between the effects of individual ability and major on earnings. Arcidiacono modeled the ability sorting in order to control for this endogeneity, and he found that differences in earnings premiums persisted. Ability sorting is beyond the scope of my paper, at least in this iteration.

Brewer, Eide, and Ehrenberg (1999) studied earnings differences by institution attended. Even after extensive controls, including sorting effects, they found convincing evidence that attending "elite" universities, especially private ones, lead to an extra earnings premium. Link (1975) also found quality of graduate schools important in predicting electrical engineers' earnings, although only for workers under age 40. That is, the effect appeared to be temporary. Craft & Baker (2003) looked at earnings of lawyers by degree, both by undergraduate major and by level of education. They only found one undergraduate major significantly associated with different earnings; lawyers who studied economics for their Bachelor's degree had higher salaries.

Card (1999) wrote a chapter in the Handbook of Labor Economics which surveyed literature on the "causal relationship between education and earnings" (p. 1802). As mentioned above, he discussed the merits of instrumental variables and extensive heterogeneity modeling, such as that in twin studies. In his review, Card concludes that OLS comes very close to correctly quantifying the average return to education. However, he posits there is a slight upward bias, probably because of unobserved ability. Card notes that "factors… that are associated with higher *returns* to education are also generally associated with higher *levels* of education" (p. 1854). In other words, the people who get the most out of further education are the ones who take advantage of it. He also pointed out that the observed return to schooling will be higher when annual or weekly pay is examined, versus hourly pay, because more education is associated with working more hours.

Song and Orazem (2005) found that students who majored in fields associated with better quantitative skills and worse verbal skills (measured by GRE scores) were more likely to start a career immediately after they finished their bachelor's degrees. Students who were very skilled at quantitative thinking (again, as measured by GRE) appeared to have a higher opportunity cost of graduate school. The authors state (p. 15):

> *Consequently, average earnings of bachelor's degree recipients overstate the opportunity cost faced by those opting to pursue advanced degrees, and so the observed premium of average earnings for postgraduate degree holders over bachelor's degree recipients understates the true returns to graduate school. Correcting for the sorting raises the estimated returns.*

Pleis et al. (2007) examined survey response rates on income questions, and their research suggested that nonresponse on income questions is correlated with lower income, although the findings were not conclusive. This is important to keep in mind, because a substantial fraction of respondents in my dataset answered every question *but* the income question. If there is a censoring issue, it could lead to biased estimates of the regressors. Finally, Griliches (1977, 1979) notes measurement errors in reported schooling would lead to a downward bias in the estimator of the effect of schooling on earnings.

Summary

While I am ostensibly the first to examine different returns to graduate school by undergraduate subject, previous work has provided me with a well-established model framework and given me several estimation concerns to carefully consider as I interpret my results. The main implications are to lean on OLS as an estimation tool and to use a model of log earnings with demographic controls.

Data

Source

The National Survey of College Graduates' stated purpose is "to provide data useful in understanding the relationship between college education and career opportunities" ("National Survey of College Graduates", 2013). It is a longitudinal biennial survey that has been conducted since the 1970s by the National Center for Science and Engineering Statistics, a division of the National Science Foundation. It "examin[es] various characteristics of college-educated individuals, including occupation, work activities, salary, the relationship of degree field and occupation, and demographic information" ("National Survey of College Graduates", 2013).

My data comes from the 2010 survey, which contains individual-level observations on calendar year 2009 data. Demographic information is available, as well as earnings data and information on each degree an individual received. A minor issue is that the survey purposely targets engineers and scientists, although not exclusively so. Engineers and a few other STEM occupations are heavily oversampled, and sample weights are given in the dataset.

Key variables

The NSCG has many variables about respondents' personal and professional backgrounds. In my regression analysis, I settled on using annual salary as the outcome variable, degree type as the covariate of interest, and controls for age, marital status, whether children live in the home, race/ethnicity, sex, and parents' education.

Salary is expressed as a natural logarithm. *Degree type* is extracted from information on an individual's highest degree(s) received, and the individual's field of study. I created a STEM/non-STEM designation by classifying seven given degree fields[1], then I combined the level of degree variable (bachelor's, master's, Ph.D., or professional school) and STEM/non-STEM variables to create dummy variables for *STEM undergraduate degree*, *STEM graduate degree*, *non-STEM undergraduate degree*, and *STEM graduate degree and STEM undergraduate degree* (an interaction term).

One issue is that there is no variable for work experience, which was commonly found to be significant in other studies (e.g., Craft, 2003; Song, 2005). I include an age variable instead. Because college/professional schooling is controlled for, this combination of education and age essentially measures the same information as work experience (Mincer, 1974). *Marital status* is simplified to married/not married, and *children living at home* is similarly classified as a binary variable. *Race* and *mother's* or *father's education* categories can be seen in Appendix I.

Sample characteristics

My sample is comprised of respondents with known earnings who reported working the week before the survey took place. Respondents ranged from 21 years old to 75 year old. The lower age limit is not fixed but reflects the fact that all respondents have a bachelor's degree; the upper limit is imposed by the NSF.

In order to better understand the data, let's first take a look at the distribution of our dependent variable (Figure 1). Below is a histogram of earnings data.

[1] *STEM fields include "computer and mathematical sciences", "biological, agricultural, and environmental life sciences", "physical and related sciences", "engineering", and "science- and engineering-related fields". Non-STEM fields are "social and related sciences" and "non-science and engineering fields".*

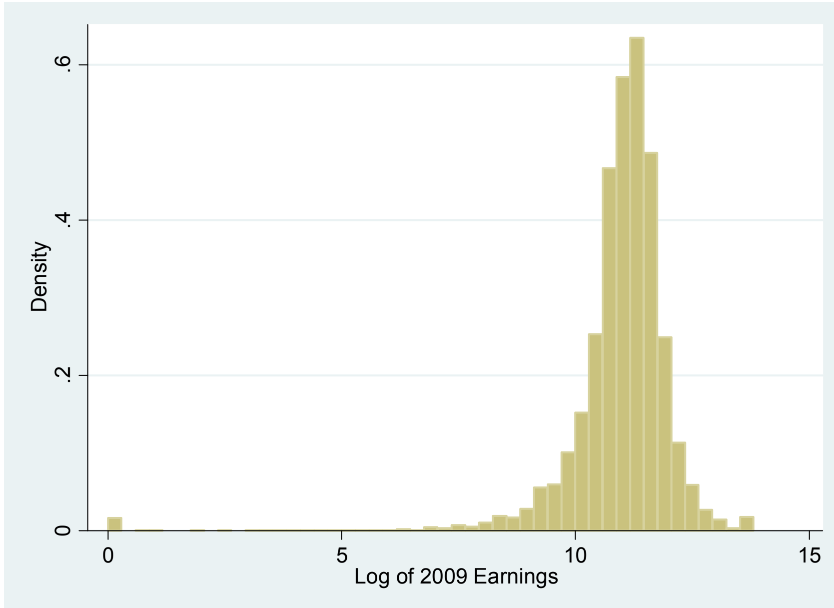


Figure 1

The mean of the earnings variable is $84,798, and the median is $65,000. The distribution does appear unusual for earnings (see Appendix III for kernel density estimates). Only four variables have the maximum value of $999,996, although many observations did not report income at all. In fact, income had higher nonresponse rates than other key variables ("National Survey of College Graduates", 2013).

Because the sample is nonrepresentative, 58% of respondents are men, 71% are married, and 71% are white. Figure 2 shows the highest degree earned of each individual. Note that all have bachelor's degrees. Appendix I has summary statistics and tabulations for all variables in the model, while Appendix II has the correlation matrix.

Highest degree type	Freq.	Percent	Cum.
Bachelors	31,715	51.47	51.47
Masters	21,214	34.43	85.90
Doctorate	4,528	7.35	93.25
Professional	4,158	6.75	100.00
Total	61,615	100.00	

Figure 2

In Figure 3 below, I show the sample means earnings by undergraduate and graduate degree types. It is apparent that earnings are higher for all individuals with STEM undergraduate degrees. However, the earnings premium of getting a STEM graduate degree is higher for individuals with non-STEM undergraduate majors—the gap closes. In fact, the mean increase in earnings with a STEM graduate degree is 62% for non-STEM undergrads, and 47% for STEM undergrads. This is preliminary evidence in favor of the original hypothesis.

	Type of undergraduate degree		
Has Non-STEM Grad. Degree?	NQ only	Q	Total
No (n=46,031)	$64,047	$87,387	$72,742
Yes (n=15,584)	$79,113	$108,255	$83,756
Has STEM Grad. Degree?			
No (n=45,463)	$67,020	$79,045	$70,069
Yes (n=16,152)	$108,580	$116,119	$114,241
Has both types of Grad. Degrees?			
No (n=59,779)	$68,676	$89,204	$75,084
Yes (n=1,836)	$98,307	$123,805	$113,507
Total (n=61,615)	$78,478	$88,775	$84,798

Figure 3

Regression Analysis

Model specification

My model is an ordinary least squares regression, as follows:

$$Y = \alpha + \beta X + \gamma_1 S_{UG} + \gamma_2 S_G + \gamma_3 NS_G + \gamma_4 S_{both} + \varepsilon$$

Y is log earnings, which is reported total earned income from all sources. Log of earnings (as opposed to the level form) was chosen because of traditional reasons: earnings takes large integer values ($1 was added so zero-valued observations were not dropped), it only takes positive values (which can lead to skew), and the log transform is less sensitive to outliers.

X contains the following: Marital status, age, gender, race, children living in household, mother's education, and father's education. The control variables are probably not exhaustive, but these demographic variables are the ones which most often appear in the literature, and should be sufficient for the task at hand. Again, I would prefer to control for experience for more explanatory power; there may actually be omitted variable bias, because experience is correlated with age. This biases the coefficient of age upwards. However, this is not a major concern, and age can be thought of as “potential” work experience (Mincer, 1974). Note that I do not control for sector in which a person is employed; the model (admittedly naively) assumes individuals maximize earnings, picking the best-paying job in any field, regardless of which field their degree is in.

The other parameters are dummy variables indicating whether the individual had a particular degree (or degree combination, in the case of S_{both}). The coefficient on S_{UG} is the premium a person expects with an undergraduate degree in a STEM field (vs. one with an undergraduate degree in a non-STEM field). The coefficient on S_G measures the earnings premium for someone who gets a STEM graduate degree. So, a person who has a non-STEM undergraduate degree and a STEM graduate degree would have expected earnings represented by:

$$E(Y|NS_{UG}, S_G) = \alpha + \gamma_2,$$

and the earnings premium for the degree is simply γ_2. The earnings of a person with a STEM undergraduate degree and a STEM graduate degree is slightly more complicated, and would be represented by:

$$E(Y|S_{UG}, S_G) = \alpha + \gamma_1 + \gamma_2 + \gamma_4.$$

Even though γ_1 (the coefficient on S_{UG}) is a part of expected earnings, it’s already accounted for, and in this case the earnings premium of the STEM graduate degree is $\gamma_2 + \gamma_4$. Hence, the interaction variable S_{both} is the key to answering the research question—if it has a significant negative coefficient, that is evidence that that people get more out of graduate STEM degrees if they do not have an undergraduate STEM degree.

Potential concerns

It is important to keep in mind a few limitations of my data when fitting the regression model. First, I needed to account for the sample weights. This was accomplished simply by declaring the weights in Stata, and then prefixing my *regress* command with *svy:*. Next, the data is top coded, so I considered a Tobit specification. This did not materially change the results, as only four observations on income took the maximum value.

As mentioned earlier, many researchers view earnings premiums as heterogeneous variables; that is, the coefficient is a random variable that differs across individuals. If that is the case, coefficient estimates will be biased toward zero. Sibling studies or twin comparisons have been employed in the past and have shown small practical differences (Card, 1999). Perhaps the best way to account for this heterogeneity is to sample the same individuals over time in a panel set. However, that would require substantial variation within individuals in order to see the idiosyncratic returns. In the end, I decided that the problem was small enough to safely ignore.

Omitted variable bias, or endogeneity caused when a causal variable is omitted and correlated with both the error and a covariate, is a somewhat more serious concern. Previous research suggests that degree type is probably correlated with underlying ability (a "selection effect"), and that could bias the results or even make estimates inefficient. Since I do not observe ability directly, I include parents' education as a set of proxy variables for ability. This should mitigate omitted variable bias, and again, thorough accounting for the problem did not negate degree premium effects in other research, as in Arcidiacono (2004).

Results

I report the relevant statistics for the key variables here (Figure 4). Full results can be seen in Appendix IV.

Key coefficients in Model 1:

Variable	Coefficient	p-value
Non-STEM grad degree	0.204	0.000
STEM grad degree	0.456	0.000
STEM undergrad degree	0.221	0.000
STEM grad & undergrad degrees	-0.125	0.000
Constant	9.394217	0.011
N = 61,615	R-squared = .0793	Prob > F = .0000

Figure 4

In order to test the hypothesis that returns to STEM graduate are greater for individuals who did not study a STEM subject as an undergraduate, I need to frame the question as a null hypothesis:

$$H_0 : \gamma_4 > 0$$

The key coefficient is γ_4 (received STEM graduate *and* undergraduate degrees) because it reflects the difference to people of different backgrounds. Remember, an individual who received a non-STEM undergraduate degree has an earnings premium of γ_2, while the earnings premium for STEM undergraduates is $\gamma_2 + \gamma_4$. If I can reject the null hypothesis that γ_4 is 0 or positive, the conclusion is that the earnings premium is indeed greater for people with non-quantitative undergraduate degrees.

In order to perform the one-sided test, the p-values are divided by two (if the coefficient has the correct sign). It is clear from the results that we can strongly reject the null, implying γ_4 is significantly negative. The interpretation is $100 * (e^{\gamma_2+\gamma_4} - 1)\%$ because it is a log-linear model. The end result is a STEM graduate degree predicts 19% higher earnings for someone with a non-STEM undergraduate degree than they would earn with a STEM undergraduate degree (a 58% increase, compared to a 39% increase).

Also significant were *male*, *married*, *kids at home*, and *age* (all positive coefficients). Certain degree indicators were positive, although most were not. Mother's and father's education variables were (separately) jointly significant. When I dropped one parental education variable, the remaining variable became much stronger. For this reason, I prefer model 3 in Appendix IV. *Father's education level* has a stronger relationship to ability, compared to that of *mother's*, due to less variation in American men's education in the previous generation or two.

Because *race* variables were insignificant, separately or jointly, I tried removing them from the model (model 2 in Appendix IV), but this did not affect other coefficients. Strictly speaking, my *both STEM degrees* variable includes anyone who received a STEM undergraduate degree, regardless of whether they also received a non-STEM undergraduate degree (3,670 individuals). This leads to overestimating the premium for undergraduate STEM degrees, and underestimating the magnitude of the *both* variable. In model 5 in Appendix IV, I see what happens when I leave out individuals with both types of undergraduate degrees. Again, the results are not changed much. The coefficient on *both* does increase in magnitude when those individuals are left out, indicating in the reduced sample we see an even larger STEM graduate degree earnings premium for people with only a non-STEM undergraduate degree, compared to STEM undergraduates.

Conclusion

My work adds to the body of literature on earnings returns to education. In summary, the National Survey of College Graduates has yielded interesting results. Regression analysis shows a large and significant difference in the STEM graduate degree premium for those who do not have STEM undergraduate degrees.

The difference I find in relative value of earning a graduate degree in a STEM field can instruct students, educators, and administrators on how to guide new Bachelor's recipients. Knowing the *relative* value of a STEM graduate degree can help individuals make informed choices about their futures. For those who are willing and able to transition from liberal arts, social science, business, etc., it can be a highly profitable endeavor to seek a graduate degree in a quantitative field.

Bibliography

Arcidiacono, Peter. “Ability Sorting and the Returns to College Major.” Journal of Econometrics 121 (1-2): 343-375, 2004.

Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. “Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings.” The Journal of Human Resources 34 (1): 104-123, 1999.

Card, David. “The Causal Effect of Education on Earnings.” Handbook of Labor Economics, Volume 3: 1801-1863, 1999.

Craft, R. Kim, and Joe G. Baker. “Do Economists Make Better Lawyers? Undergraduate Degree Field and Lawyer Earnings.” Journal of Economic Education 34 (3): 263-281, 2003.

"Educational Attainment." - U.S. Census Bureau. Web. Accessed 05 Dec. 2013. (Author not identified) <http://www.census.gov/hhes/socdemo/education/index.html>.

Griliches, Zvi. "Estimating the returns to schooling: some econometric problems." Econometrica 45:1-22, 1977.

Griliches, Zvi. "Sibling models and data in economics: beginnings of a survey." Journal of Political Economy 87: S37-S65, 1979.

Heckman, James J. and Solomon Potachek, "Empirical evidence on the functional form of the earnings-schooling relationship." Journal of the American Statistical Association 69: 350-354, 1974.

Hunt, Shane J., "Income Determinants for College Graduates and the Return to Educational Investment." Yale Economic Essays 3, Fall Issue: 305-57, 1963.

Link, Charles R., “Graduate education, school quality, experience, student ability, and earnings.” The Journal of Business, Vol. 48, No. 4: 477-491, 1975.

Mincer, Jacob. *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research, 1974.

Pleis, J.R., J.M. Dahlhamer, and P.S. Meyer, “Unfolding the answers? Income brackets and income nonresponse in the National Health Interview Survey (NHIS).” ASA Proceedings of the 2006 Joint Statistical Meetings, 2007.

Song, Moohoun and Peter Orazem, “Returns to graduate and professional education: the roles of mathematical and verbal skills by major.” Working paper #05028, October 2005.

Appendices

Appendix I –Summary statistics & tabulations

Summary of earnings data

Percentiles			
1%	1600		
5%	12000		
10%	22000	Obs	61615
25%	42000	Sum of Wgt.	61615
50%	68000	Mean	84798.38
		Std. Dev.	86950.99
75%	100000		
90%	150000	Variance	7.56e+09
95%	200000	Skewness	5.299698
99%	430000	Kurtosis	44.54918

Earnings by STEM graduate degree, Y/N

	mean	p50	sd
N	78876.56	63336	82395.93
Y	101466.5	83000	96724.81
Total	84798.38	68000	86950.99

Father's education	Freq.	Percent	Cum.
Less than HS	8,224	13.35	13.35
HS	14,793	24.01	37.36
Some College	10,674	17.32	54.68
Bachelors	13,960	22.66	77.34
Masters	7,136	11.58	88.92
Professional	3,339	5.42	94.34
Doctorate	2,796	4.54	98.88
N/A	693	1.12	100.00
Total	61,615	100.00	

Mother's education	Freq.	Percent	Cum.
Less than HS	8,479	13.76	13.76
HS	19,303	31.33	45.09
Some College	12,912	20.96	66.05
Bachelors	12,693	20.60	86.65
Masters	6,097	9.90	96.54
Professional	879	1.43	97.97
Doctorate	834	1.35	99.32
N/A	418	0.68	100.00
Total	61,615	100.00	

Race	Freq.	Percent	Cum.
Asian ONLY	10,144	16.46	16.46
American Indian/Alaska Native ONLY	333	0.54	17.00
Black ONLY	5,833	9.47	26.47
White ONLY	43,465	70.54	97.01
Native Hawaiian/ Other Pacific Islander ONLY	287	0.47	97.48
Multiple Race	1,553	2.52	100.00
Total	61,615	100.00	

Gender	Freq.	Percent	Cum.
F	25,915	42.06	42.06
M	35,700	57.94	100.00
Total	61,615	100.00	

Children in household?	Freq.	Percent	Cum.
N	31,367	50.91	50.91
Y	30,248	49.09	100.00
Total	61,615	100.00	

Married?	Freq.	Percent	Cum.
N	17,986	29.19	29.19
Y	43,629	70.81	100.00
Total	61,615	100.00	

Non-STEM G	Freq.	Percent	Cum.
N	46,031	74.71	74.71
Y	15,584	25.29	100.00
Total	61,615	100.00	

STEM G deg.	Freq.	Percent	Cum.
N	45,463	73.79	73.79
Y	16,152	26.21	100.00
Total	61,615	100.00	

STEM UG deg.	Freq.	Percent	Cum.
N	23,794	38.62	38.62
Y	37,821	61.38	100.00
Total	61,615	100.00	

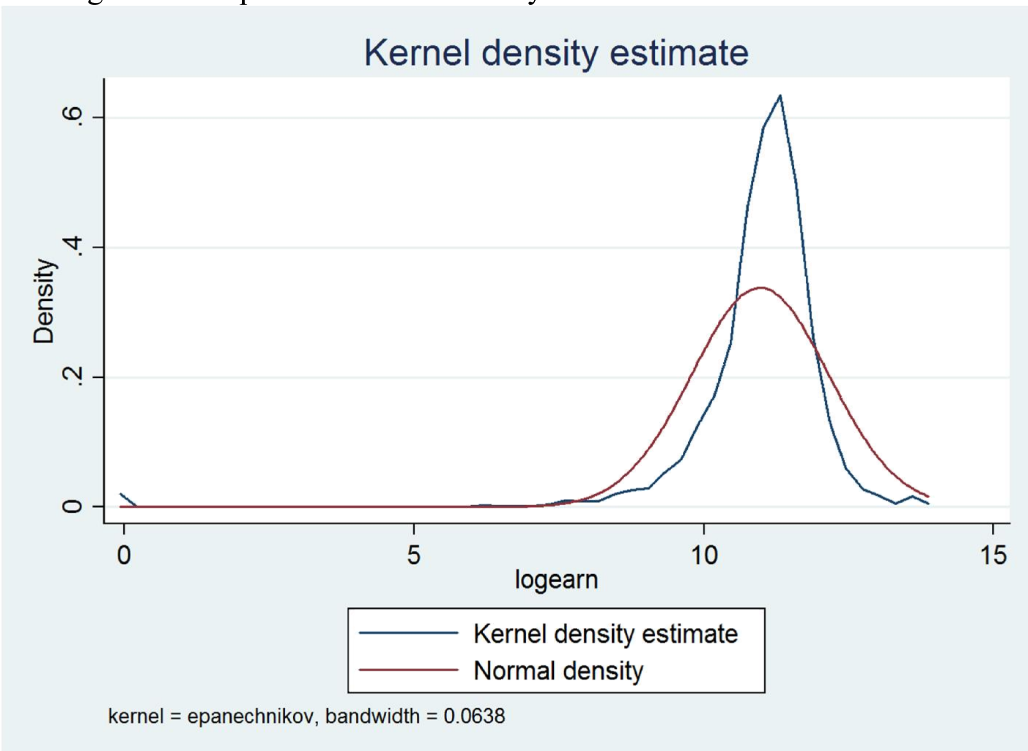
Highest degree type	Freq.	Percent	Cum.
Bachelors	31,715	51.47	51.47
Masters	21,214	34.43	85.90
Doctorate	4,528	7.35	93.25
Professional	4,158	6.75	100.00
Total	61,615	100.00	

Appendix II – Correlation matrix

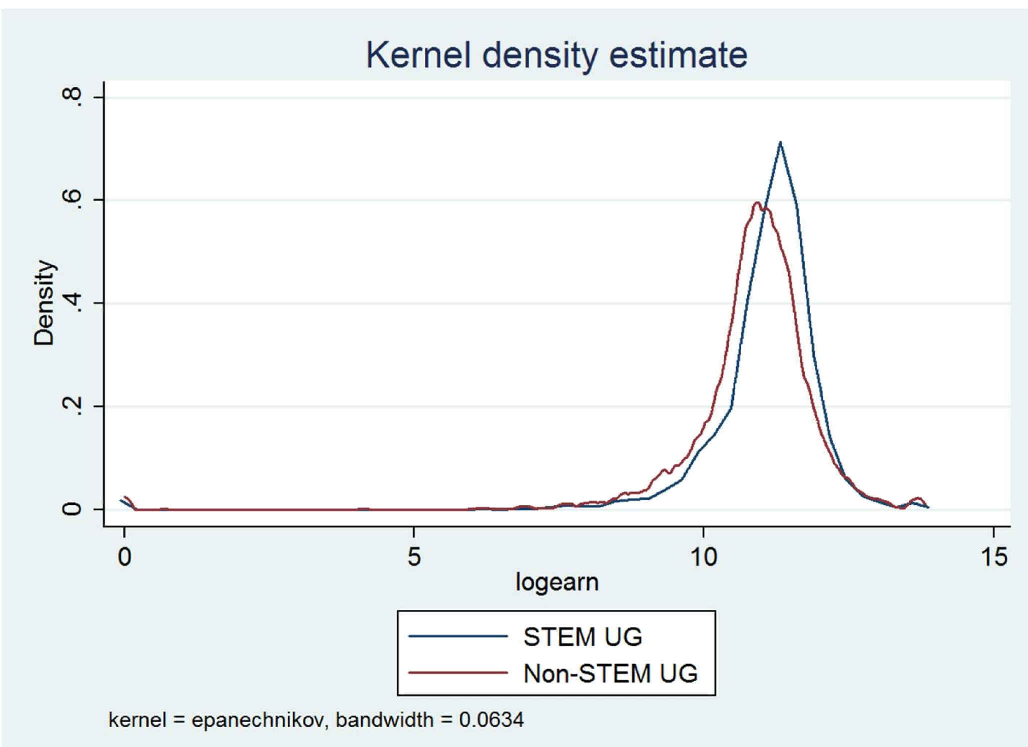
	earn	G_NQ	G_Q	UG_Q	racem	gender~c	chlvin~c	age	marind~c
earn	1.0000								
G_NQ	0.0513	1.0000							
	0.0000								
G_Q	0.1143	-0.1850	1.0000						
	0.0000	0.0000							
UG_Q	0.0609	-0.3330	0.2334	1.0000					
	0.0000	0.0000	0.0000						
racem	-0.0195	0.0573	-0.1396	-0.0973	1.0000				
	0.0000	0.0000	0.0000	0.0000					
gender_enc	0.1993	-0.0868	0.0455	0.1747	-0.0064	1.0000			
	0.0000	0.0000	0.0000	0.0000	0.0952				
chlvin_enc	0.0922	-0.0231	0.0321	0.0534	-0.0589	0.0291	1.0000		
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
age	0.1460	0.0947	0.0506	-0.0099	0.0631	0.1144	-0.0528	1.0000	
	0.0000	0.0000	0.0000	0.0104	0.0000	0.0000	0.0000		
marind_enc	0.1345	-0.0037	0.0758	0.0793	-0.0257	0.1437	0.3890	0.2107	1.0000
	0.0000	0.3331	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Appendix III – Earnings density estimates

Earnings data compared to normal density:



Earnings data by undergraduate degree type:



Appendix IV – Regressions

	Dependent variable: log earnings				
Model:	1	2	3	4	5
Variables					
Has non-STEM graduate degree	0.204***	0.206***	0.204***	0.206***	0.205***
	-6.39	-6.42	-6.39	-6.52	-6.18
Has STEM graduate degree	0.456***	0.452***	0.457***	0.460***	0.457***
	-9.98	-10.03	-10.02	-10.07	-9.97
Has STEM undergraduate degree	0.221***	0.217***	0.221***	0.221***	0.237***
	-9.09	-9.15	-9.12	-9.1	-9.63
Has STEM grad. & undergrad. degrees	-0.125*	-0.124*	-0.125*	-0.126*	-0.141**
	(-2.53)	(-2.52)	(-2.53)	(-2.55)	(-2.83)
Race†: American Indian/Alaska native	-0.426		-0.418	-0.432	-0.463
	(-0.87)		(-0.86)	(-0.89)	(-0.89)
Race: Black	0.0298		0.0329	0.0135	0.0278
	-0.43		-0.48	-0.2	-0.39
Race: White	0.0561		0.0656	0.0507	0.0629
	-1.19		-1.46	-1.08	-1.29
Race: Hawaiian/Pacific Islander	-0.614		-0.608	-0.62	-0.64
	(-0.90)		(-0.89)	(-0.92)	(-0.85)
Race: Multiple races	-0.0271		-0.0206	-0.0368	-0.0234
	(-0.16)		(-0.12)	(-0.22)	(-0.14)
Gender: Male	0.469***	0.469***	0.470***	0.472***	0.473***
	-17.82	-17.57	-17.92	-18.02	-17.36
Children living at home?	0.172***	0.168***	0.172***	0.174***	0.173***
	-6.41	-6.24	-6.49	-6.53	-6.23
Age	0.00447***	0.00468***	0.00452***	0.00434***	0.00451***
	-3.6	-3.71	-3.67	-3.48	-3.5
Married?	0.0882**	0.0903**	0.0866**	0.0906**	0.0807*
	-2.81	-2.98	-2.76	-2.88	-2.48
Mother's ed.‡: High school diploma	0.0738	0.0869		0.112*	0.0711
	-1.37	-1.67		-2.51	-1.28
Mother's ed.: Some college	0.0218	0.036		0.0584	0.0167
	-0.36	-0.61		-1.2	-0.26
Mother's ed.: Bachelor’s degree	0.0492	0.0579		0.106*	0.0503
	-0.8	-0.95		-2.17	-0.79
Mother's ed.: Master’s degree	0.0433	0.0572		0.0979	0.0403
	-0.56	-0.77		-1.47	-0.5
Mother's ed.: Professional degree	0.213*	0.221*		0.275***	0.212*
	-2.33	-2.44		-3.36	-2.22
Mother's ed.: Doctorate	0.0389	0.0514		0.0958	0.0264

	-0.4	-0.55		-1.12	-0.27
Mother's ed. N/A:	-0.177	-0.179		-0.185	-0.171
	(-1.47)	(-1.48)		(-1.75)	(-1.39)
Father's ed. ‡: High school diploma	0.0892	0.0899	0.117**		0.087
	-1.72	-1.75	-2.72		-1.61
Father's ed.: Some college	0.0263	0.0276	0.0422		0.0217
	-0.44	-0.47	-0.9		-0.35
Father's ed.: Bachelor's degree	0.112	0.113	0.132**		0.111
	-1.88	-1.94	-2.81		-1.79
Father's ed.: Master's degree	0.0627	0.0647	0.0809		0.0622
	-0.87	-0.9	-1.29		-0.83
Father's ed.: Professional degree	0.139	0.14	0.168*		0.127
	-1.71	-1.73	-2.36		-1.49
Father's ed.: Doctorate	0.106	0.108	0.127*		0.111
	-1.51	-1.56	-2.24		-1.51
Father's ed.: N/A	0.00433	-0.00132	-0.0437		-0.00115
	-0.04	(-0.01)	(-0.46)		(-0.01)
Constant	9.155***	9.181***	9.174***	9.187***	9.157***
	-106.88	-110.96	-112.19	-105.03	-103.47
Observations	61615	61615	61615	61615	57945
R-squared	0.079	0.078	0.079	0.078	0.079

t statistics in parentheses
** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*
†: omitted category is Asian
‡: omitted category is "Some high school"

Appendix V – Blinder-Oaxaca decomposition

```
Blinder-Oaxaca decomposition
Number of strata   =         1         Number of obs      =       61615
Number of PSUs     =     61615         Population size    =       40165866
            Design df          =       61614
1: UG_Q = 0
2: UG_Q = 1

Linearized
logearn       Coef.   Std. Err. t      P>t     [95% Conf.  Interval]

Differential
Prediction_1    10.68772   .0186243    573.86 0.000     10.65122  10.72422
Prediction_2     11.0244   .0105337    1046.59      0.000     11.00376  11.04505
Difference     -.3366831   .0213968    -15.74 0.000    -.3786208  -.2947453

Decomposition
Endowments   -.1281353   .0088231      -14.52 0.000    -.1454285  -.110842
Coefficients   -.1783496   .0217712    -8.19  0.000    -.2210212  -.135678
Interaction   -.0301982    .015067     -2.00  0.045    -.0597296  -.0006669
```