QM222 Project: Returns to College

Deyuan(Elliot) Wang :wdy@bu.edu
Yifei(Danny) Zhao: yifeiz@bu.edu
QM222 Section B1
April.21st.2021

Preface:
All values from the regressions' coefficients are listed with a unit measurement of dollars/$.
i.e. 28954 = $28954

Introduction
According to the United States Census Bureau, from 2010 to 2019, the percentage of people who are age 25 and older with a bachelor's degree or higher jumped from 29.9% to 36.0%. It has been embedded into the public's mind that a good education is the main variable of the formula to success. In the subject, we used a series of data that collects information of randomly selected 3,221 employed individuals and ran regressions based on the variables that are considered important to demonstrate whether a college education is worth the investment.

Regression A analysis
Regression A (Table 1) only contains one variable, which measures the relationship between college graduates and individual earnings. From this regression, it is stated that at a 95% confidence level, college graduates make 28,954 dollars more than non-college graduates, and the average individual earning for non-college graduates is 40,220 dollars. Those patterns were fully proved to be identical to the statistical result tested within the previous data analysis (Table 2) from Part A, in which the average earning for a nongraduate was 22,639 dollars, whereas the average earnings for a college graduate was 68,073 dollars. However, this regression is limited and suffers from omitted variable bias that there are other variables that bias the college graduate's coefficient. In the following analysis, more variables will be analyzed for lessening the bias.

Regression B analysis
Regression B (Table 1) contains two more variables compared to Regression A, gender and rural residence. The regression reflected the phenomenon that a male on average makes $19,057 more than a female, and a person who resides in a city averagely makes 6,489 more than those who live in a rural area. Both corresponding standard errors suggested 95% confidence levels. Interestingly, the coefficient for college graduates increased from $28,954 to $30,714 according to the previous regression. Intuitively, it might be the fact that gender has some negative bias towards college graduates since a female makes significantly less than a male even though she has gone to college and more females tend to earn college degrees than males. To prove the theory, two additional regressions (Regression F and Regression G) were tested with different explanatory variables: gender and rural residence. It has been confirmed that the effect on college graduate's coefficient is caused by gender variance since the rural regression did not have a significant impact on the coefficient, while in the gender regression, the coefficient of college graduates was largely increased.

Regression C analysis

Regression C (Table 1) contains one more variable compared to Regression B, IQ percentile. From this regression, the coefficient for gender and rural residence marginally changed, yet the coefficient for college graduates was decreased by a significant amount. It is very likely that the newly added variable, IQ percentile, has lowered the coefficient of college graduates. It is at a 95% confidence level that for every unit of increase of IQ percentile, the individual earnings will increase by 287 dollars. The omitted variable bias was lessened that the coefficient for college graduates decreased from 30,714 dollars to 22,604 dollars since the IQ percentile has a positive effect on the coefficient of college graduates, that people with higher IQ percentile are more likely to receive a college degree.

Regression D analysis

Regression D (Table 1) is the full model according to the variables provided. It contains all the variables in Regression C along with working hours, marital status, and parent college education. From the regression, it is examined that all of the three added variables have applied some degree of positive effects on the coefficient of college graduate. According to the test, it is at a 95% confidence level that every unit of increase of working hours increases the individual earnings by $14.15 dollars, a married status increases the earnings by $9,843 dollars, and a person whose parent's education is equal or above college degree has an increase in earnings by $2,948 dollars. The omitted variable bias was further lessened that the coefficient for college graduates decreased from $22,603 dollars to $20,625 dollars.

Regression E analysis

Despite college graduates being the main factor of earning variance, it is also worth mentioning how interesting variables interact differently when gender is considered. To further understand the interrelations among the gender variance and its causation to the earnings gap, another model(Figure 3) was developed from Regression C to include and examine two extra variables, marital status of males and females. To explain how the effect of the marital status attribute to the gender pay gap, it is worth noticing that on average males make $9, 157 more than females do. Based on this presumption, the marital status even amplified the earnings for males in a more significant way. From regression E, it is a 95% confidence level that a married male has comparatively 18,264 dollars more earnings than those who are not married. On the other hand, married females merely get 2,500 dollars increases in earnings, not a very significant amount compared to males.

Regression F:

Throughout the finding, it was noticed that one variable from Regression C, IQ percentile, contributed to decrease the coefficient of college graduates by 8,110, which is around 26.4%. Thus, an additional significant regression was run with two more variables, CollegeGrad*IQ and NonCollegeGrad*IQ. Surprisingly, the 95% coefficient for the CollegeGrad*IQ was 523, and the 95% coefficient for the latter variable was 172. Simply, when holding the IQ percentile constant, a college graduate is likely to earn significantly more than a nongraduate. In conclusion , IQ percentile has a significant positive effect on the coefficient of college graduates.

Conclusion:

Does a college degree really worth the investment? The experiment to answer such a question was indeed complicated. Initially, the research started to test from comprehensive perspectives, showing the correlation between individual earnings and college degree. From regression A to D, the research found out how other explanatory variables like rural residence, gender, area, work hours, IQ percentile, marital status, and parent education are considered to apply omitted variable bias to the coefficient of the college graduate variable. By adding multiple variables, it demonstrates that the (full) model successfully decreased the biased effect from the simple model, in which the only variable considered is college graduate. With the significance of college graduates remaining constant, other variables were proved to affect the individual earnings by variance of degrees. The regression's outcome, without exception, still satisfies the result of positive correlation of college degrees with the increase of individual earnings. As the significance of college graduates remains, other variables . Secondly**,** there are confounding factors, gender and IQ percentile, that also indirectly affect the coefficients of both college graduate and individual earnings. According to a series of tested regressions, it turns out that IQ percentile also applies a positive effect on the average earnings. When a college degree is added, this pattern is even more amplified.

In conclusion, it is no surprise that investment for college education can still be a viable choice to enhance one's career profitability in the current society. As individual's career ages grow, the effect of a college degree will become incrementally significant.

Appendix:

Table1: Regression Table

	A	B	C	D
College Graduates	28954.44***	30714.53***	22603.70***	20625.89***
	(882.75)	(1326.00)	(1488.66)	(1433.57)
Gender (Male=1)		19057.06***	18279.43***	12740.78***
		(1306.60)	(1283.87)	(1255.15)
Area (Rural=1)		-6489.07***	-6232.67***	-8056.93***
		(1815.38)	(1781.34)	(1698.27)
IQ Percentile			286.87***	245.12***
			(25.59)	(25.15)
Working Hours				14.15***
				(0.82)
Marital Status (Married=1)				9843.86***
				(1237.21)
Parent Education (Mother_College=1)				2948.97**
				(1318.26)
Intercept	40220.64***	31255.08***	20292.27***	-11153.52***
# of observations	3221	3221	3221	3221
SEE	38091.63	36829.74	36136.12	34256.12
Adjusted R-squared	0.12359	0.18044	0.21102	0.29098

Table2:

Descriptive statistics of annual earnings by college-graduate and employment status:

earnings with no college degree(employed)		*earnings with college degree(employed)*	
Mean	40223.98	Mean	69175.08
Standard Error	610.37	Standard Error	1353.53
Median	35000.00	Median	56000.00
Mode	40000.00	Mode	235884.00
Standard Deviation	26331.10	Standard Deviation	49897.20
Sample Variance	693326987.11	Sample Variance	2489730496.52
Range	235484.00	Range	233984.00
Minimum	400.00	Minimum	1900.00
Maximum	235884.00	Maximum	235884.00
Sum	74856829.00	Sum	94008933.00
Count	1861.00	Count	1359.00
25th percentile	28000.00	25th percentile	40000.00
75th percentile	65000.00	75th percentile	80000.00
Confidence Level(95.0%)	1197.09	Confidence Level(95.0%)	2655.23

The difference in two mean earnings is identical to the coefficient variable of college_graduate

Figure1: Regression B coefficients(excluding Rural =1)

Regression Statistics	
Multiple R	0.42183935
R Square	0.17794844
Adjusted R Square	0.17743753
Standard Error	36897.0691
Observations	3221

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	30098.9121	1101.16388	27.3337264	4.161E-148	27939.8585	32257.9657
Male = 1	19094.8883	1308.94912	14.587953	1.0065E-46	16528.4298	21661.3467
college_graduate=1	31081.3246	1324.44862	23.4673691	1.498E-112	28484.4762	33678.1729

Figure2: Regression B coefficients(excluding Male =1)

Regression Statistics	
Multiple R	0.35644975
R Square	0.12705643
Adjusted R Square	0.12651389
Standard Error	38022.0411
Observations	3221

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	41394.2982	940.25052	44.0247544	0	39550.7477	43237.8488
rural=1	-6703.5408	1874.09309	-3.5769519	0.00035274	-10378.078	-3029.0037
college_graduate=1	28579.8788	1360.57135	21.0057921	6.5441E-92	25912.2046	31247.553

Figure3: Regression E coefficients(including married male/female =1)

Regression Statistics	
Multiple R	0.48567037
R Square	0.23587571
Adjusted R Square	0.23444922
Standard Error	35595.445
Observations	3221

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	20437.7087	1691.237882	12.08446718	6.5278E-33	17121.6946	23753.7228
IQ_percentile	268.101857	25.32527368	10.58633601	9.1135E-26	218.446533	317.757181
rural=1	-7501.992	1763.224458	-4.254700502	2.1531E-05	-10959.15	-4044.8336
male	9157.01927	1902.56692	4.812981439	1.5555E-06	5426.65182	12887.3867
married male	18264.6785	1836.668594	9.94446063	5.704E-23	14663.5181	21865.839
married female	2497.51899	1770.909353	1.410303123	0.15854697	-974.70716	5969.74515
college_graduate=1	21830.0738	1469.637282	14.85405555	2.5323E-48	18948.5525	24711.5951

Figure4: Regression F coefficients

Regression Statistics	
Multiple R	0.41246063
R Square	0.17012377
Adjusted R Square	0.16934987
Standard Error	37078.0165

Observations	3221

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	33417.4093	1575.874881	21.20562343	1.6036E-93	30327.5888	36507.2298
IQ*college grad	523.620614	42.21084566	12.40488328	1.4976E-34	440.857738	606.38349
IQ*non grad	172.250238	33.44631494	5.150051302	2.7608E-07	106.671992	237.828483
college_graduate=1	403.769972	3408.443152	0.118461701	0.90570924	-6279.1702	7086.71017