A picture containing person, crowd

Description automatically generated

**Is College A Good Investment?**

A Multiple Regression Model Analysis

Chen Dong

cdong27@bu.edu

**Introduction**

Following the explosive growth of contemporary technologies and massive improvements of education, more and more people choose to attend college in order to become more competitive candidates in the job market. However, as time passes, employers begin to doubt the necessity of college degrees as a proof of workers’ capability. A recent article on *The New York Times* raises a question: is college a good investment? To address this question, this report adopts multiple regression analysis to figure out if college degrees can boost people’s annual earnings as the public expects. In response to *The New York Times*’s question, our results show that although college indeed has a positive impact on annual earnings, when more variables are considered in the model, this impact gradually shrinks.

**Analysis**

**Regressions of college graduation - earnings relationship**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| College Graduate | $48,391\*\*\*  (24.88) | $33,747\*\*\*  (15.07) | 17.52\*\*\*  (37.52) |
| IQ |  | $836\*\*\*  (12.46) |  |
| Intercept | $32,569\*\*\*  (31.56) | $(47,194)\*\*\*  (-7.28) | 95.42\*\*\*  (385.14) |
| # Observations | 3704 | 3704 | 3704 |
| SEE | 53240.22 | 52164.37 | 12.78 |
| Adjusted R2 | 14% | 18% | 28% |

t-statistics in parentheses; \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

The first regression introduces the college graduates-earnings relationship. As we all know, most people go to college because they believe getting into universities can boost their earnings, so we will compare these two variables first using the simple regression model. It has three parts: the limited model, the full model, and the background. The limited model, regression 1, only compares earnings and the college graduation status. The intercept means that non-college graduates will earn $32,569, while college graduates on average will earn $48,391 more than non-college graduates. It indicates a strong gap between them. However, it is important for us to take IQ into account because it can be very influential. The full model, regression 2, explains the relationship among earnings, IQ, and college graduates. $(47,194) means that if a student is not college graduate with an IQ of zero, he will have to give out this amount of money. This does not make sense because most people have IQ above 70. $836 means that for every extra 1 IQ a person has, his earning is expected to increase by $836. If this person is college graduate, he should earn $33,747 more than non-college graduate. The background, regression 3, explains that for a non-college graduate, the IQ is expected to be 95.42, and a college graduate has 17.52 points higher IQ. By applying the omitted-variables-bias formula, C1 = B1 + A1 \* B2, we have 48393 = 33,747+ 17.52\*836. So, there is a difference about $15,000. That is to say, the coefficient from the limited regression will decrease by approximately 15,000 when we include IQ which indicate that along with the add-in of new variables, the college degree effect is likely to decrease.

**Multiple regression of college graduation, gender and rural - earnings relationship**

|  |  |
| --- | --- |
|  | 4 |
| College Graduate | $48,232\*\*\*  (25.67) |
| Gender | $29,498\*\*\*  (17.49) |
| Rural | $3,002\*  (1.67) |
| Intercept | $16,961\*\*\*  (9.98) |
| # Observations | 3704 |
| SEE | 51155.88 |
| Adjusted R2 | 21% |

t-statistics in parentheses; \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

In order to further explain the influence of college attendance and IQ regarding earnings, it is crucial to include gender and rural areas. Gender is such a sensitive but consequential factor that it must be considered. Researchers are debating about whether women are discriminated in workplace, which results in lower salaries. Also, the variable of rural area has to be included, because most people believe that urban areas provide more job opportunities with high wage rates. As regression 4 above shows, if a female who is not a college graduate and lives in rural area, she is expected to earn $16,961. If she is a college graduate, she should earn $48,232 more. If she does not live-in rural area, she will earn $3,002 more. If the person is male, he will earn $29,498 more than a female. However, this regression cannot explain too much because it does not account for IQ.

**Regressions of college graduation, IQ, rural and gender – earnings and gender-hours relationship**

|  |  |  |
| --- | --- | --- |
|  | 5 | 6 |
| College Graduate | $34,607\*\*\* (15.98) |  |
| Gender | $28,397\*\*\*  (17.14) | 7.01\*\*\*  (7.62) |
| Rural | $3,233\*  (1.83) |  |
| IQ | $776\*\*\*  (12.00) |  |
| Intercept | $(56,717)\*\*\*  (-8.92) | 34.41\*\*\*  (55.01) |
| # Observations | 3704 | 3704 |
| SEE | 50194.79 | 27.91 |
| Adjusted R2 | 24% | 2% |

t-statistics in parentheses; \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

As a result, regression 5 includes IQ in the regression. A non-college female with zero IQ who lives in rural area is expected to have negative $56,717 annual earnings (this does not make sense because the universal average IQ is 100). If she is a college graduate, she should earn $34,607 more. If she does not live in rural area, she will earn $3,233 more. With all else equal, a male will earn $28,397 more than a female. Moreover, for each extra point of IQ he has, he is expected to earn $776 more. In other words, just being “talented” (people with IQ above 120) can boost the salary by $15,520, which means that even though a person chooses not to go to college, he is still going to earn a lot simply due to his high IQ. Also, the calculated omitted-variable bias is about $14,000 (as regression 2 is a limited model while regression 3 is the full model). That means by only adding one variable IQ, the value of college attendance can decrease by $14,000. After that, living in urban areas can boost people’s earnings by $3,000 which is not a lot (however, it does not necessarily indicate a true improvement of life quality as urban areas usually have higher price levels). Besides that, being a male does not mean that he will definitely earn more because, according to the regression 6, males typically work 7 hours more than women weekly. However, it is reasonable to conclude that as more different categories are added in, the college attendance effect on earnings gradually decreases.

**Multiple regression of working hours and IQ – earnings relationship**

|  |  |
| --- | --- |
|  | 9 |
| College Graduate | $(74,116)\*\*\*  (-4.19) |
| IQ\*Graduate | $1,507\*\*\*  (10.55) |
| IQ\*Non-Graduate | $534\*\*\*  (7.32) |
| Hours | $897\*\*\*  (14.42) |
| Hours Squared | $(5)\*\*\*  (-9.45) |
| Northeast | $7,699\*\*\*  (3.27) |
| Intercept | $(41,576)\*\*\*  (-5.92) |
| # Observations | 3704 |
| SEE | 50249.77 |
| Adjusted R2 | 24% |

t-statistics in parentheses; \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

This regression model is chosen with the consideration about three generally recognized determinants of people’s annual earnings, including college degrees, hours worked per week, and the degree of development of the area where people work. In multiple regression model D, the adjusted R square is 24%, indicating that 24% of the data can be explained by this model. According to the coefficients of the intercept and all X variables, the model can be expressed as:

Earnings=-41,576-74,116\*College Graduate+1,507\*IQ\*Graduate+534\*IQ\*Non-Graduate+897\*Hours-5\*Hours Squared+7,699\*Northeast+e

In this model, -74,116 indicates that college graduates on average earn $(74,116) less annual income than non-graduates; 1507 is the extra annual earning per 1 point of IQ for college graduates, while 534 is that for non-graduates; 897 and 5 are the additional annual earnings per hour or per hour squared worked per week, separately; finally, if a person works in the northeast, he or she will be able to earn $7,699 more annual income than others on average. All the coefficients are statistically significant, as the t stats for each of them are greater than 2. The intercept suggests that a non-college graduate who has a 0 IQ and works 0 hour per week in an area other than northeast earns $(41,576) annual income. This might sound strange in real world, because it is almost impossible for a person to have a 0 IQ or for the government to collect money from a person who does not work (except for winning a lottery). However, for a non-college graduate with average IQ and hours worked per week and does not work in the northeast, the model shows annual earnings of $37,844.67. This is a more acceptable number of annual earnings. This outcome can be explained by the large coefficient of the IQ of non-graduates and annual earnings’ non-linear relationship with hours worked per week. For a non-college person with average IQ of 100, the boost of earnings, $534,00, already exceeds the intercept, while the hours worked per week shows a non-linear pattern similar to the left side of a downward parabola, which increases rapidly at first and then flattens out. The relationship suggests that as a person works more hours per week, his annual earnings increases rapidly at first, and the rate of increase gradually decreases to near zero. Another noticeable unusual value is the coefficient of being a college graduate. The model shows that attending college can result in $(74,116) fewer annual earnings than non-college people, which seems to indicate that college is an unworthy investment. However, this difference can be explained by the coefficient of IQ\*Graduate. The boost of income by one extra point of IQ for graduate people reaches approximately 3 times of that for non-graduate people. Holding all else equal, as for a person with average IQ of 100, the boost of income if he or she is a college graduate will be $150,700, while that number will be only 53,400 if the person is a non-college people. The difference, $97,300, exceeds the negative impact of the variable “College Graduate,” so eventually college proves itself to be a valuable investment for people who want to receive higher annual incomes.

**Regressions of handedness-earnings relationship**

|  |  |  |
| --- | --- | --- |
|  | 7 | 8 |
| College Graduate |  | $34,605\*\*\*  (15.98) |
| Right-Handed | $555  (0.20) | $1,254  (0.52) |
| Rural |  | $(3,247)\*  (-1.84) |
| IQ |  | $775\*\*\*  (11.99) |
| Gender |  | $28,431\*\*\*  (17.14) |
| Intercept | $45,714\*\*\*  (17.63) | $(54,533)\*\*\*  (-8.26) |
| # Observations | 3704 | 3704 |
| SEE | 57519.10 | 50199.77 |
| Adjusted R2 | 0% | 24% |

t-statistics in parentheses; \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

Another interesting topic is “Do lefties usually earn more?” Many people claim that lefties are smarter as people like Leonardo Da Vinci, Winston Churchill, and Aristotle are all left-handed. Two regressions are made to investigate this topic. Regression 7 indicates that right-handed people earn $555 higher annual incomes. Considering that lefties earn $45,714 on average, $555 is relatively neglectable, while its t statistics also suggest that lefties and righties earn the same number of salaries statistically. However, it is important to consider other variables, so the regression 8 contains rural areas, IQ, gender, and college degree variables. In regression 8, right-handed people are expected to earn $1254 more. Similarly, because the t statistics for right-handed is 0.52, which is smaller than 2, regression 8 suggests the same conclusion as regression 7 did before: there is no difference between two types of handedness statistically.

**Conclusion**

In conclusion, college attendance has a positive effect on the earnings overall. However, IQ also contributes to the earning to a large extent. Also, it is reasonable to assume that while more explanatory variables are included (in other words, the value of R squared increases), college degree’s influence on earnings gradually decreases. The quote from New York Times is still debatable as college degrees still have an impact on people’s earnings. Moreover, people should not simply regard college as a boost for earnings because colleges provides people with plenty of opportunities to meet talented students, reputable professors, and fantastic recruiters. These resources cannot be measured in monetary values in short terms. However, this report should only be viewed as a reference, because the data set is collected in the form of survey. Therefore, this data set only includes the responses of people who are willing to participate, which is not randomized, and the self-selection bias applies to all regressions conducted above. For instance, people who cannot afford college but have high IQ are more willing to participate in the survey to demonstrate that even they do not attend college, they can still have desirable annual earnings.