Do Computer Science graduates earn higher yearly incomes compared to other undergraduate majors?

Minh T. Nguyen

DePauw University

Submitted in partial fulfillment of the requirements for Econ 385: Regression and Simulation

## 

## 

## Acknowledgements

I would like to express my appreciation to Professor Humberto Barreto of the Economics and Management Department at DePauw University. I would also like to sincerely thank you to my classmates—Joel Magnabosco, Tyler Myers, and Aidan Shea for their helpful feedbacks during the peer reviewing process.

## Abstract

This paper uses the Current Population Survey (CPS) data to examine whether college graduates with a bachelor’s degree in Computer Science earn higher annual income than graduates with other majors or people without a college degree. This paper will estimate a variety of variables to mitigate the effect of unobserved variables that might be correlated with the yearly income of the observations. The result suggests that Computer Science and Engineering majors make more than other majors in general. This study also discusses the results related to the gender and racial pay gap, as well as the effect of work experience and educational attainment.

## Introduction

In recent years, the demand for technical workers, especially in software engineering has always been on the rise, to the extent that US companies need to hire foreign workers to adapt to their business growth. By numbers, there were more than 200,000 H1-B visa holders in 2002, accompanied by more than 20$ million worth of scholarship money given yearly to undergraduate computer science, engineering, and math majors, according to data from the National Science Foundation (Mervis, 2000). The number of foreign workers authorized to work in the US increased to nearly 600,000, according to the 2019 USCIS H1-B Population Survey.

With the provided context, the demand for workers with at least a bachelor’s degree is on the rise, so the goal of this study is to observe how an education in the mentioned field will pay off in the future.

## Literature Review and Theoretical Analysis

This literature review will discuss the body of research related to the research question of this paper: What is the monetary return on education for a computer science graduate, compared to other majors? To do so, I will first discuss the recent publications on this subject and briefly present results from previous studies. Then, this review will introduce the independent variables that can influence the answer to the research question, explaining why these factors are relevant and their characteristics. Then, I will discuss the theoretical framework applied to the study to develop the regression model.

As mentioned above, the return on education has long been a topic of interest for researchers. In his study on such topics, Arcidiacono (2004) points out that the return on majoring in natural science and business is generous. Arcidiacono also found a strong relationship between undergraduate GPA and yearly earnings. For all majors in general, going from 2.5 to 3.0 will increase yearly earnings by 5%, especially for business majors the increase is 13%. However, regarding natural science majors, he notes that majors such as biology might have higher grades in college, but not necessarily be well-compensated in the workforce, which leads to depressing coefficients on yearly earnings (Arcidiacono, 2004).

The racial income gap is also relevant in this context. Recent studies suggest that the educational attainment gap between Blacks and Whites is narrowing, especially since the rate of college attendance for African Americans increases more than that of White-Americans–4.4% to 22.5% compared to 11.3% to 32.8% in the same period (McClough and Benedict, 2017). However, this increase does not necessarily strongly affect narrowing the race pay gap. The researchers also conclude that African American college students, on average, graduate with academic majors associated with lower-paying occupations. These researchers also point out the underlying reason for this conclusion, which is that Black high school graduates are academically less prepared than Whites coming to college. That said, the number of Black male students switching from rigorous majors such as natural sciences, engineering, economics, or business to less academically demanding majors like social sciences and humanities is 54% compared to 8% of White males. However, they suggest that students with higher ability will be able to complete rigorous coursework associated with higher pay, regardless of race or gender (McClough and Benedict, 2017).

On the other hand, it is evident that gender does play a role in determining income after graduation. Joy (2003) points out that academic-related variables such as major, grades, or institution attended do influence the gender salary gap, but the effect of the said variables is not as substantial as labor market-related variables. However, Joy also suggests that academic majors and the labor market are only a part of the story, as 75% of the gender wage gap remained unexplained. This proves that the gender pay gap remains debatable, as many previous studies failed to produce a consistent answer to this problem. Krueger (1993) points out that while the gender pay gap is narrowing, for the most part, researchers have not been able to fully explain the earnings gap, regardless of controlling for factors such as labor supply, work experience, or training (Daymont and Andrisani, 1984). Therefore, gender needs to be included as a control to improve the regression coefficients.

Echoing the above discussion on the effect of race and gender on annual income after college, years of educational attainment are added for its relevance. On the other hand, years of work experience are derived using this variable, and I will explain by what formula experience is calculated as I continue my analysis.

To further improve the regression results, work experience is also added as a control independent variable. Since the relationship between income and work experience is non-linear, a quadratic transformation will be applied, meaning that the squared of experience is added to the model equation.

Since the purpose of this study revolves around observing the monetary return on education, I must take into consideration the time value of money, as the investment in education yields returns in the future. Then, I need to consider the present value (in US Dollars) of the return on education after graduation, in this case, is the income of graduates. The earnings function framework is the chosen solution to this problem. This theoretical framework was developed by Mincer (1958, p. 283) and explains that the distribution of personal income, which is our interest, only approximates normality and is symmetric when it is transformed into a logarithmic distribution. That is, personal income differentials are described via a random shock process, which “generates a log-normal distribution if applied to the logarithms of income rather than to income itself” (Mincer, 1958). Then, the income differentials will be interpreted as a percentage difference, rather than an absolute value of income.

## Empirical Analysis

### A. The Data

To answer the research question, I will be using sample data from the 2020 American Community Survey (2020 ACS) provided by the Integrated Public Use Microdata Series (IPUMS USA). I will be extracting 1% of the dataset size, and after cleaning up and the data generation process has been completed, the number of observations being used for the regression model is 41,384 (**Ruggles et al., 2022)**.

**1.Data Generation Process and Model Specification**

**Model Specification:**

As previously discussed, I will be using the earnings function framework to develop a regression model. A natural log transformation will be applied to the dependent variable, which details will be explained below. For the independent variable, the earnings function will produce results which cannot be interpreted as absolute values in US Dollars. Rather, the coefficients are interpreted as a percentage change in annual income after adjusting for the properties of natural logs by using

##### **Dependent variable:**

I choose to measure yearly income (in US Dollars) over hourly pay for several reasons. The first reason is to be consistent with other studies on annual income. More importantly, using yearly income allows us to control full-time workers. An hourly-based wage fails to indicate whether a person is working full-time or not. I decide that the threshold for full-time workers is an annual income of over $20,000. Therefore, any observations with INCWAGE less than 20000 will be omitted. Another reason is that low INCWAGE values when being applied to logarithmic functional form will yield erroneous results which can potentially distort the regression results. Additionally, college graduates will likely be paid on a salaried basis instead of an hourly basis. That is, if a graduate works as a software developer, contractor, engineer, or in managerial roles, he or she will be paid on a salary basis, which includes a fixed amount of base salary, adding bonuses, commissions, stock options, equity, etc. No amount of overtime work will be recorded on a salary basis. Therefore, using yearly total compensation is a more liable measure since it corrects for fluctuations in the annual income of a worker.

##### **Independent variables:**

For this paper, different majors will be aggregated into several categories--Computer Science (CS), Natural Sciences (NaturalScience), Social Sciences (SocialScience), Engineering (Engineering), Business (Business), Arts (Arts), Other Majors (OtherMajors), and No College Degrees (NoDeg). Then, the unrestricted regression model will use NoDeg as a categorical variable base case to which other aggregated major groups can be compared to. Also, the question of whether a graduate double majored in college is also taken as a control to improve the regression result. Since the research question is examining the annual income of CS majors, any person who has Computer Science as their second degree will also be counted towards as a CS major

Variables related to race and gender will be used as control variables. Years of education attainment (EducYears) ranges from 9 to 18; grade 1 to 9 is aggregated to 9 years of education and having more than 18 years of education is categorized as 18. Years of experience (Exp) will be calculated by subtracting years of education attainment (EducYears) and 6 (from 0 to 6 years before the first year of schooling) from AGE. The assumption is that a person will have their first year of working experience after completing their education, which starts at 6 years of age.

**2.Summary Statistics**

This part of the study will present a summary of the dataset. Firstly, a summary statistic from the independent variables is presented in Table I.

Table I: Summary Statistic of all independent variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean | Std. Dev. | Min | Max |
| CS | 0.0201383 | 0.1404749 | 0 | 1 |
| Engineering | 0.0508897 | 0.219775 | 0 | 1 |
| NaturalSci~e | 0.0916014 | 0.2884659 | 0 | 1 |
| SocialScie~e | 0.1647568 | 0.3709654 | 0 | 1 |
| Business | 0.0937772 | 0.291522 | 0 | 1 |
| Arts | 0.0177449 | 0.1320244 | 0 | 1 |
| OtherMajors | 0.0179625 | 0.1328166 | 0 | 1 |
| NoDeg | 0.5446524 | 0.4980082 | 0 | 1 |
| EducYears | 14.42044 | 2.439814 | 9 | 18 |
| Exp | 24.38819 | 11.98118 | 1 | 50 |
| N (Observations) | 41,364 |  |  |  |

The table demonstrates that there are about 2% of observations major in Computer Science, while Engineering is 5%. Social Science is the most popular major group in the dataset at 16%, while Business and Natural Science are roughly the same at 9%. However, more than half of the sample does not have a college education, at 54%, as the average year of education is around 14.4, while that of working experience is 24.4.

For the following section, I will present data on CS-related topics, and the focus is examining the average income of CS and non-CS majors in relation to years of education, race, and gender. However, the averages of income do have their limitations. For example, presenting the summary of income after graduation of CS and non-CS graduates fails to control the effect of gender, race, and other academic majors. Since the averages fails to account for other controlling factors, these results only serve as a brief overview of the dataset, while no conclusion is being made from these data. Having said that, the regression results will explain the relationship between different variables in their contribution to yearly income.

1. Income after graduation of CS and non-CS graduates

Table II: Means of wage and salary income on CS and EducYears

|  |  |  |  |
| --- | --- | --- | --- |
|  | CS(1=Yes) |  |  |
| EducYears | 0 | 1 | Total |
| 9 | 42997.585 | . | 42997.585 |
| 10 | 46118.033 | . | 46118.033 |
| 11 | 45250.321 | . | 45250.321 |
| 12 | 51691.47 | . | 51691.47 |
| 13 | 58779.428 | . | 58779.428 |
| 14 | 58763.69 | . | 58763.69 |
| 16 | 85201.854 | 105227.77 | 86224.722 |
| 18 | 113696.12 | 135525 | 114445.03 |
| Total | 73108.712 | 114829.77 | 73948.902 |

Table II demonstrates the average income of CS graduates compared to those without a CS degree (both attained at least 16 years of education), and others without a college degree. Initially, the dataset is consistent with the assumption that more education means more income. We can see an income gap of roughly $20,000 *($105,227.77- $85,201.854)* between CS and non-CS graduates. Additionally, a college degree is still worth investing in—college grads make $35,000 *($85,201.854- $51,691.47)* more than those without a degree—12 years of education at max.

1. Racial pay gap in Computer Science

Table III: Means of wage and salary income on CS and White

|  |  |  |  |
| --- | --- | --- | --- |
|  | White (1=Yes) | |  |
| CS | 0 | 1 | Total |
| 0 | 64541.925 | 76698.652 | 73108.712 |
| 1 | 115349.51 | 114523.28 | 114829.77 |
| Total | 65820.598 | 77380.083 | 73948.902 |

The table above shows the average income of CS and non-CS graduates by race. While there is a big income disparity between CS and non-CS majors regardless of race—from about $50,000 *($115,349.51-$64,541.925)* to $39,000 *($114,523.28-$76,698.652)*, the pay between White and non-White Computer Science graduates is effectively the same.

1. Gender income disparity in Computer Science

Table IV: Means of wage and salary income on CS and Male

|  |  |  |  |
| --- | --- | --- | --- |
|  | Male (1 = Yes) | |  |
| CS (1 =Yes) | 0 | 1 | Total |
| 0 | 62750.029 | 82067.23 | 73160.494 |
| 1 | 93256 | 122059.66 | 115513.38 |
| Total | 63032.612 | 83126.126 | 73948.902 |

Table IV illustrates the gender income gap among Computer Science graduates. While it is evident that graduating with a CS degree leads to more annual income, there exists a noticeable pay gap between males and females, as male makes substantially more that their counterpart regardless of academic major. The regression results will contribute to the discussion of the gender income differential by controlling for other factors such as years of work experience, education, and race.

### B. Presentation and Interpretation of Results

##### **Regression Results**

This portion of the study will report the Ordinary Least Squared regression result of regression the natural log of yearly income on Majors dummy, Double Majors, Race, Ethnicity, Gender, Education and Work Experience. There are 3 regression models being displayed in the table below. The first model is the restricted regression of the natural log of income on CS. However, since Model 1 only regress CS major on lnINCWAGE, the base case is non-CS majors. Model (2) and (3) will gradually include other control statistics such as Majors, Race, Ethnicity, Gender, Education and Work Experience (NoDeg as the base case). Model 3 is the unrestricted regression of this paper.

Table V: OLS Regression Results (Dependent Variable: lnINCWAGE)

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
|  | lnINCWAGE | lnINCWAGE | lnINCWAGE |
| CS | 0.472\*\*\* | 0.320\*\*\* | 0.233\*\*\* |
|  | (0.0281) | (0.0299) | (0.0292) |
|  |  |  |  |
| NaturalScience |  | 0.147\*\*\* | 0.155\*\*\* |
|  |  | (0.0206) | (0.0202) |
|  |  |  |  |
| SocialScience |  | 0.00713 | 0.0217 |
|  |  | (0.0192) | (0.0186) |
|  |  |  |  |
| Engineering |  | 0.350\*\*\* | 0.252\*\*\* |
|  |  | (0.0248) | (0.0244) |
|  |  |  |  |
| Business |  | 0.200\*\*\* | 0.170\*\*\* |
|  |  | (0.0214) | (0.0207) |
|  |  |  |  |
| Arts |  | -0.0627 | -0.0717\* |
|  |  | (0.0360) | (0.0349) |
|  |  |  |  |
| OtherMajors |  | 0.00144 | -0.00153 |
|  |  | (0.0325) | (0.0311) |
|  |  |  |  |
| DoubleMajor |  | 0.0580\*\* | 0.0565\*\* |
|  |  | (0.0188) | (0.0183) |
|  |  |  |  |
| EducYears |  | 0.0948\*\*\* | 0.0977\*\*\* |
|  |  | (0.00340) | (0.00343) |
|  |  |  |  |
| Exp |  | 0.0360\*\*\* | 0.0354\*\*\* |
|  |  | (0.00138) | (0.00134) |
|  |  |  |  |
| Exp2 |  | -0.000550\*\*\* | -0.000542\*\*\* |
|  |  | (0.0000280) | (0.0000272) |
|  |  |  |  |
|  |  |  |  |
| female |  |  | -0.246\*\*\* |
|  |  |  | (0.00756) |
|  |  |  |  |
|  |  |  |  |
| black/african american |  |  | -0.175\*\*\* |
|  |  |  | (0.0124) |
|  |  |  |  |
| american indian or alaska native |  |  | -0.119\*\*\* |
|  |  |  | (0.0333) |
|  |  |  |  |
| chinese |  |  | 0.0334 |
|  |  |  | (0.0317) |
|  |  |  |  |
| japanese |  |  | 0.0536 |
|  |  |  | (0.0534) |
|  |  |  |  |
| other asian or pacific islander |  |  | -0.0285 |
|  |  |  | (0.0176) |
|  |  |  |  |
| other race, nec |  |  | -0.0312 |
|  |  |  | (0.0193) |
|  |  |  |  |
| two major races |  |  | -0.0386\* |
|  |  |  | (0.0150) |
|  |  |  |  |
| three or more major races |  |  | -0.00970 |
|  |  |  | (0.0555) |
|  |  |  |  |
|  |  |  |  |
| mexican |  |  | -0.0873\*\*\* |
|  |  |  | (0.0159) |
|  |  |  |  |
| puerto rican |  |  | -0.0732\* |
|  |  |  | (0.0316) |
|  |  |  |  |
| cuban |  |  | -0.176\*\*\* |
|  |  |  | (0.0503) |
|  |  |  |  |
| other |  |  | -0.101\*\*\* |
|  |  |  | (0.0210) |
|  |  |  |  |
| Constant | 10.92\*\*\* | 9.052\*\*\* | 9.180\*\*\* |
|  | (0.00435) | (0.0452) | (0.0463) |
| Observations | 41364 | 41364 | 41364 |
| *R*2 | 0.011 | 0.225 | 0.272 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

##### **Interpretation of Results**

Since the earnings function framework is being applied to the dataset, the error term of the regression model should be considered. The regression model is influenced by omitted variable bias and these biases are embedded into the error term of the model. The source of omitted variable bias and other factors, as well as the measures taken to mitigate these errors, will also be discussed as I continue my analysis. Adding more control variables produced mixed regression results. While some majors improve their annual income by holding other variables constant, that is not the case for others.

Model 1 points out that CS majors make about 60.3±0.03% (e0.472±0.028) more than non-CS majors in general—which includes those without a college degree. Controlling other academic major groups, education, and experience, the income of CS majors is about 37.7±0.03% (e0.320±0.0299) more than that with no college education. The regression model continues to report a slight decrease, as CS majors make 26.23±0.029%(e0.0233±0.0292) more, as race, ethnicity, and gender are included. This result suggests that race, ethnicity, and gender have an impact on the annual income of CS majors, as mentioned in previous studies. Specifically, the slight decrease can be explained by using White, Not Hispanic, and Male as the base case for constructing dummy control variables. Meaning, CS majors who are not White, Hispanic, or Female make slightly less than their counterparts. These findings can be contributed to the gender and racial income gap discussed in the literature review. Additionally, we can see that education and experience contribute to the decreasing coefficients of CS. While I have not found any studies that suggest a decrease in income as work experience increases, the results suggest that this decline can be credited to the non-linear characteristics of experience.

The annual income of other majors comparing with having no college education is also worth noticing, as they can be compared to that of CS majors. Model 2 suggests that Engineering majors make approximately 42±0.025% (e0.35±0.0248) more that of not having a college degree, holding all other variables constant. For model 3, Engineering remains the highest paying major, while Computer Science comes second at 26.23±0.03% (e0.233±0.0292) more that having no college education, holding all other variables constant.

For Natural Science majors, the number is roughly 16.76±0.02% (e0.155±0.0202), which is noticeably lower than CS and Engineering given that they are commonly assumed to be academically rigorous majors. Arcidiacono (2004) mentioned that this can be contributed to the fact that majors such as biology are underpaid. On the other hand, having a business degree pays 18.53±0.02% (e0.17±0.0207) than no college education, holding all other variables constant. That is, while the numbers for Natural Science and Business majors are comparable, graduating with a degree in one of these fields will earn you less money than Computer Science or Engineering.

Social Science, Arts, and other major groups are associated with lower-paying fields, and the regression reports inconclusive numbers. The unrestricted regression model reports a coefficient of 0.0216±0.018 for Social Science. Therefore, [-0.0144, 0.0576] is the 95% confidence interval for this coefficient. Since this interval covers zero, we cannot make any conclusion about the income of Social Science majors. The same argument can be made with OtherMajors.

The regression results also suggest that one year of education will increase annual income by 10.26±0.00343% (e0.0977±0.00343), while the case for work experience is 3.6±0.00134% (e0.0354±0.00134). For both models 2 and 3, these numbers remain effectively similar.

##### **Omitted Variable Bias, Heteroskedasticity, and other limitations**

I acknowledge that the regression results above are hugely affected by omitted variable bias. The unrestricted regression model (Model 3) only accounts for roughly 27.2% of the variance in the dataset (R2 = 0.272). The positive takeaway is that the R2 value increases as more control variables are added. We can see an increase in R2 from 1.1% (R2 = 0.011) for Model 1 to 22.5% (R2 = 0.225). for model 2. That is, the control variables chosen to include do influence omitted variable bias, but they can do so far as explaining one-fourth of the variance. That said, this regression model fails to include other variables that can influence annual income. There are several reasons for this issue. While some of the variables are not recorded by the IPUMS USA dataset, there are factors that cannot be measured both qualitatively and quantitatively. For example, there are no specific metrics for measure chance, luck, motivation, or parental influence. Other factors that might affect annual income are unemployment, parental education, etc.

Heteroskedasticity in the dataset is also a concern. To tackle this problem, I report the Robust Standard Error (Robust SE), instead of the normal SE. By doing so, the hypothesis test results turn out that a major part of the independent variables reports statistically significant results at p-values of 0.01% or lower. However, there are certain variables such as Chinese or Japanese that report higher p-value or are statistically insignificant. Since they are control statistics, they only have a slight impact on the interpretation of the regression results. On the other hand, the high p-value (more than 5%) for the coefficient of SocialScience and OtherMajors implies that assuming these coefficients do not affect lnINCWAGE, it is likely that we can get this result. Hence, we cannot make any conclusion regarding the impact of SocialScience and OtherMajors.

Additionally, limitations from the IPUMS USA dataset are worth mentioning. The data collection process for the CPS is vulnerable to biases since not every person is equally likely to be interviewed and sampled in IPUMS. For example, certain racial groups or geographical locations are more likely to be surveyed than others. To tackle this issue, I also include PERWT, which is the ACS variable for the probability of being selected. Per the description of the PERWT variable provided by IPUMS USA, it represents “how many persons in the U.S. population are represented by a given person in an IPUMS sample” (IPUMS USA, 2022)

## Conclusion

In conclusion, I found out that CS and Engineering are among the highest-paid occupations. However, Natural Science is underpaid compared to CS and Engineering, consistent with previous studies pointing out this issue (Arcidiacono, 2004). Continuing the discussion raised by previous studies, the gender pay gap is unexplained by factors such as majors, education, work experience, and race. However, the coefficients on annual income of Social Science majors remains inconclusive. These results suggest that the demand for Computer Science-related discipline is still on the rise, and the overall return on investment for education is generous.

## Works Cited

Arcidiacono, Peter. “Ability sorting and the returns to college major”. Journal of Econometrics, 2004, vol. 121, issue 1-2, 343-375. <https://doi.org/10.1016/j.jeconom.2003.10.010>. Accessed 28 Nov. 2022.

Daymont, Thomas N., and Paul J. Andrisani. “Job Preferences, College Major, and the Gender Gap in Earnings.” *The Journal of Human Resources*, vol. 19, no. 3, 1984, pp. 408–28. *JSTOR*, https://doi.org/10.2307/145880. Accessed 18 Nov. 2022.

Joy, Lois. “Salaries of Recent Male and Female College Graduates: Educational and Labor Market Effects.” *Industrial and Labor Relations Review*, vol. 56, no. 4, 2003, pp. 606–21. *JSTOR*, https://doi.org/10.2307/3590959. Accessed 28 Nov. 2022.

Krueger, Alan B. “How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989.” *The Quarterly Journal of Economics*, vol. 108, no. 1, 1993, pp. 33–60. *JSTOR*, https://doi.org/10.2307/2118494. Accessed 18 Nov. 2022.

McClough, David, and Mary Ellen Benedict. “Not All Education Is Created Equal: How Choice of Academic Major Affects the Racial Salary Gap.” *The American Economist*, vol. 62, no. 2, 2017, pp. 184–205. *JSTOR*, https://www.jstor.org/stable/26754414. Accessed 18 Nov. 2022.

Mervis, Jeffrey. “Demand for Tech Workers Benefits Undergraduates.” *Science*, vol. 288, no. 5463, 2000, pp. 40–41. *JSTOR*, http://www.jstor.org/stable/3074923. Accessed 17 Nov. 2022.

Unites States Citizenship and Immigration Services, “H-1B Authorized-to-Work Population Estimate”, Office of Policy & Strategy Policy Research Division, 2019, <https://www.uscis.gov/sites/default/files/document/reports/USCIS%20H-1B%20Authorized%20to%20Work%20Report.pdf>. Accessed 28 Nov. 2022.

**University of Minnesota, Minnesota Population Center. “PERWT Description”.**

<https://usa.ipums.org/usa-action/variables/PERWT#description_section>. Accessed 8 Dec. 2022.

**Steven Ruggles, Sarah Flood, Ronald Goeken, Megan Schouweiler and Matthew Sobek.***IPUMS****USA: Version 12.0* [dataset]. Minneapolis, MN:**IPUMS**, 2022.**<https://doi.org/10.18128/D010.V12.0>