# Estimating Premiums for Education in the California State: Does Gender Matter?

Data from 2016 American Community Survey of California

Qianhui Guo, Jillian Pflugrath Bass, Ying Xue
ECON 5300

Professor Hiedemann

Feb. 27th, 2019

# 1. Introduction

The pursuit of higher education is a substantial investment of time and money. According to the National Center for Education Statistics (NCES), “for the 2015-16 academic year, annual current dollar prices for undergraduate tuition, fees, room, and board were estimated to be $16,757 at public institutions, $43,065 at private nonprofit institutions, and $23,776 at private for-profit institutions”. After adjusting for inflation, those prices are up 34 percent, 26 percent, and 16 percent, respectively, from the 2005-06 academic year. People may wonder, now, more than ever, whether furthering education is a worthwhile investment. Part of making an informed choice is knowing how earnings vary based on educational attainment.

Another concern when discussing earnings is the gender pay gap. Women may question whether they will enjoy the same premiums for their investment in higher education. According to the US Department of Labor’s Women’s Bureau, “Women who worked full-time, year-round in 2014 earned 79% on average, of men's median annual earnings.” Over the last few decades, the wage gap has gotten smaller as a greater portion of women are graduating college and participating in the labor force. Women now outpace men in educational attainment. In 1975, the portion of women in the labor force aged 25 and older with a college degree was only 14%. By 2015, that portion increased to 41%. Also during this time, women’s participation in the labor force increased from 46.3% in 1975 to 56.7% in 2015. But even with these improvements, the Department of Labor reports that “progress has stalled”. Although women are now graduating college at a higher rate than men, are women seeing proportional gains in earnings for their increased educational attainment? T o better understand the overall gender pay gap, it’s important to know how education is impacting women’s earnings.

We focus our research of these topics on two questions: 1) In the state of California, how do earnings vary by educational attainment? 2) Does the premium for higher education vary by gender? By estimating a multi-linear regression earnings equation, we take into account a variety of socioeconomic and demographic variables that impact earnings and examine annual earnings per educational attainment group. We then use that same model specification to examine women’s earnings and men’s earnings separately to see if the premiums for higher educational attainment vary by gender.

Based on our regression results, earnings do vary by educational attainment: on average, the higher the educational attainment, the higher the earnings. And, while that overall pattern is consistent, men and women have different premiums for higher educational attainment.

## 2. Econometric Models and Estimation Methods

Ordinary least squares (OLS) estimators are used to estimate our multi-linear regression earnings equation. We use dummy variables for the following categorical variables: race, gender, marital status, and educational attainment. The excluded variable for race is other, for gender is female, for marital status is never married, and for education is high school diploma. Age as well as age-squared are also included in the model. Age is an important factor in explaining earnings because it is a good proxy for experience. Age-squared captures the parabolic relationship between age and earnings. Throughout an individual’s career, earnings tend to increase at a decreasing rate until maximum earnings are reached, and then earnings decrease at an increasing rate. Marital status is included, as a married person tends to earn more than an unmarried person. And race is included to control for the race pay gap in the United States.

The variances of residuals against the explanatory variables may not be homogeneous, so we test for heteroskedasticity using the BPG test. Because the BPG test is statistically significant at the 1% level, we reject the null hypothesis that the variance of the residuals is constant and infer that the model has heteroskedasticity. To address the heteroskedasticity, the dependent variable earnings is converted to the natural log of earnings. We perform the BPG test again to test the log earnings estimated equation for heteroskedasticity. The second BPG test is statistically significant at the 1% level, so we again reject the null hypothesis that the variance of the residuals is constant and infer that the model has heteroskedasticity. Because there is still heteroskedasticity, and the standard errors may be overestimated or underestimated, we calculate robust standard errors.

## 3. Data

We obtain data from the 2016 American Community Survey and focus on the state of California, amounting to 376,035 individual observations. To measure annual earnings, we aggregate income from employment and self-employment earnings for each individual. Earnings are in 2015 US Dollars. We restrict our sample to full-time year-round earners who work at least 30 hours per week. Thirty hours is our minimum because the IRS classifies a full-time worker as “an employee employed on average at least 30 hours of service per week, or 130 hours of service per month.” We notice that self-employment earnings have a large range: from -$5,800 to $853,000 with a median of $43,000. Thus, we restrict our sample to individuals who report positive earnings, and we eliminate the top 1% and bottom 1% of income earners in our sample to remove extreme outliers.

For educational attainment we use the school attainment variable. Respondents reported the highest level of education they have completed. Although the survey includes more delineations, we split the educational attainment data into nine groups: no school, some school but no high school diplom a, high school diploma or equivalent, some college but no degree, associate’s degree, bachelor’s degree, master’s degree, professional degree beyond bachelor’s, and doctorate degree. The professional degree category includes medical doctors and lawyers, while the doctorate degree category includes PhD and EdD. While professional degree is listed before doctorate degree in the survey and in our tables, it is not necessarily a lower level of educational attainment, as the category includes types of doctorate degrees. For marital status, we group the data into five categories: married, widowed, divorced, separated, and never married. For ethnicity and race, we also group the data into five categories: White, Black, Asian, Indian or Alaska Native, and other races.

We restrict our sample to individuals age 17-64 to focus on the working-age population. We set 64 as the maximum age, because according to the U.S. Census Bureau, the average retirement age for Californians in 2016 was 64.

After making the aforementioned adjustments, our sample size is 144,748. As shown in Table 1, the mean earnings for our selected sample is $64,075 ranging from $1,750 to $504,000, which indicates a large range across full-time workers in California. The average age is 41.8 years with 60% of the sample being male and 60% of the sample being married; with pressure to provide for a family, married individuals may be more motivated to pursue full-time employment, work harder, and avoid layoff compared to the unmarried individuals. The population of females in the full-time job market is approximately 20 percentage points less than that of males; more women may be leaving their job or working part time to care for children. There are various educational attainment levels in the sample, with 20% of the sample reporting high school diploma, 20% reporting some college no degree, and 20% reporting bachelor’s degree.

## 4. Empirical Result

As shown in Table 2 and Table 3, we considered four model specifications. The dependent variable is the log of earnings, and the independent variables include educational attainment, gender, age, marital status, and race. Table 2 shows the variability of earnings based on educational attainment, and Table 3 exhibits the variability of earnings based on educational attainment by gender.

In the first specification of Table 2, we regress the log of earnings on the education attainments and gender. The coefficient of educational attainment for less than high school diploma is negative and statistically significant at the 1% level. By contrast the coefficients for all educational attainment groups above high school diploma are positive and statistically significant at the 1% level. Increases in educational attainment lead to an increase on average in earnings. A full-time, year-round worker with a professional degree is expected to earn the most among all education attainment groups, earning 285.7% more than those with only a high school diploma and 28.1% more than those with a master's degree.These results are statistically significant at the 1% level. The regression results also show the gender pay gap. Controlling for educational attainment, male workers earn on average 32.3% more than female workers. This result is statistically significant at the 1% level. The adjusted R-squared in the model is 0.237, showing that the variation of educational attainment and gender only explain 23.7% of the variation in annual earnings.

In the second specification of Table 2, we add age and age-squared, marital status and race as independent variables to improve the fit of the model and address omitted variable bias. After adding these variables, the coefficients for each educational attainment category reduce, suggesting educational attainment is capturing the partial effects of marital status, age, and race in our first model. The coefficients for age, married, divorce d, Asian, White, and Indian and Alaska Native people are all statistically significant at the 1% level.

After controlling for race, marital status, age in the second specification of Table 2, and correcting for heteroskedasticity with robust standard errors, the coefficients for gender and for each educational attainment group remain statistically significant at the 1% significance level but the magnitude decrease slightly. Controlling for age, gender, marital status, and race, on average the earnings of those with a professional degree or doctorate degree are still the two highest, earning 215.2% and 189.8% respectively more than the one with high school diploma. These relationships are both statistically significant at the 1% level. A full-time, year-round worker with a master’s degree also has a large premium compared to workers with only a high school diploma, earning 146.0% more on average. The impact of education attainment on earnings is material. For example, if a person with a high school diploma earns $35,000 annually, holding age, gender, marital status, and race constant, a person with a master’s degree would be expected to earn on average $86,100 annually. After controlling for age, marital status and race in the multiple regression, the adjusted R-squared increases from 0.237 to 0.353, showing that 35.3% of the variation in earnings is explained by the explanatory variables in our second model.

Table 3 presents estimates of log earnings regressions for males and females respectively. By separating the sample into males and females and running our estimation model on both groups, we are able to compare the estimated coefficients within each gender group. Holding marital status, age, and race constant, the premiums for higher education do vary by gender. For example, men with a master’s degree earn on average 27.8% more than men with a bachelor’s degree; for women the same premium is only 22.6%. Men with a professional degree earn on average 23.0% more than men with a master’s degree, while women with a professional degree earn on average 33.0% more than women with a master’s degree. The adjusted R-squared for the male and finale models is 0.369 and 0.323 respectively, suggesting that the factors explaining earnings may be different for males and females.

## 5. Conclusion

In the state of California, on average higher educational attainment leads to increased earnings. The investment in education is on average worthwhile. Within a few years, the increase in earnings associated with further education is on average more than the cost incurred for that education.

The premiums for education attainment do vary by gender. If the premiums were exactly the same, we would not have different coefficients in the male and female models. However, the coefficients for males and females in Table 3 cannot be compared to make an claims about the gender pay gap and education. Future research could be done using interaction terms to examine the gender earning gaps per educational attainment group.

Because our sample is restricted to survey respondents in the state of California, our results cannot necessarily explain earnings patterns across the United States. While California has a large population and a very large economy supported by many industries, it may not be representative of the larger population of earners in the United States. These results could potentially be very different in states where the economy relies heavily male-dominated, unskilled labor.

Variables omitted from our estimated earnings equation pose limitations to our findings but also point to opportunities for future research. Controlling for the presence of children under 5 and children over 5 could help to better understand women's earnings. Generally the presence of young children has a greater impact on women's earnings than on men's. Similarly, we use the same model to estimate women's earnings and men's earnings, but the difference in adjusted R-squared for these two estimated equations suggests that variables impacting men's earnings may be different than tho se impacting women's earnings. Future research could specify different estimated earnings equations for men and women.

Our estimated equation may also be more accurate if we were able to control for the industry or job type, as it has a large impact on earnings for both men and women. We do not include the industry information from the survey data as we find its categories too broad for inferring meaning. Specific industry information may explain a large part of the variation in earning, help us to better understand the gender pay gap, and open up questions for future research. Are women choosing to use their education in less lucrative industries or lower paying positions? Or, are those more lucrative industries and higher paying positions inherently sexist?

Finally, while our earnings equation does control for race, further research could be done to examine the intersectionality of gender, race, and education and its effects on earnings.

## References

"Fast Facts: Tuition costs of colleges and universities," National Center for Education Statistics, https://nces.ed.gov/fastfacts/display.asp?id=76.

"Breaking Down the Gender Wage Gap," Women's Bureau, United States Department of Labor, https://www.dol.gov/wb/media/gender_wage_gap.pdf.

"Identifying Full-Time Employees," Internal Revenue Service, July 2, 2018, https://www.irs.gov/affordable-care-act/employers/identifying-full-time-employees

"Average Retirement Age in the California," U.S. Census Bureau data 2016, https://www.thebalance.com/average-retirement-age-in-the-united-states-2388864

Table 1. Descriptive statistics

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| earnings | 144,748 | 64,075.4 | 72,089.1 | 1,750 | 24,000 | 80,000 | 504,000 |
| age | 144,748 | 41.8 | 12.3 | 17 | 31 | 52 | 64 |
| Indian.Alaska | 144,748 | 0.02 | 0.1 | 0 | 0 | 0 | 1 |
| Asian | 144,748 | 0.2 | 0.4 | 0 | 0 | 0 | 1 |
| Black | 144,748 | 0.1 | 0.2 | 0 | 0 | 0 | 1 |
| White | 144,748 | 0.6 | 0.5 | 0 | 0 | 1 | 1 |
| male | 144,748 | 0.6 | 0.5 | 0 | 0 | 1 | 1 |
| Married | 144,748 | 0.6 | 0.5 | 0 | 0 | 1 | 1 |
| Widowed | 144,748 | 0.01 | 0.1 | 0 | 0 | 0 | 1 |
| Divorced | 144,748 | 0.1 | 0.3 | 0 | 0 | 0 | 1 |
| Separated | 144,748 | 0.02 | 0.1 | 0 | 0 | 0 | 1 |
| Non_school | 144,748 | 0.02 | 0.1 | 0 | 0 | 0 | 1 |
| No_HS_diploma | 144,748 | 0.1 | 0.3 | 0 | 0 | 0 | 1 |
| No_college_degree | 144,748 | 0.2 | 0.4 | 0 | 0 | 0 | 1 |
| associate | 144,748 | 0.1 | 0.3 | 0 | 0 | 0 | 1 |
| bachelor | 144,748 | 0.2 | 0.4 | 0 | 0 | 0 | 1 |
| Master | 144,748 | 0.1 | 0.3 | 0 | 0 | 0 | 1 |
| Professional | 144,748 | 0.03 | 0.2 | 0 | 0 | 0 | 1 |
| Doctor | 144,748 | 0.02 | 0.1 | 0 | 0 | 0 | 1 |

Table 2. Estimated ln(Earnings) Equations

| | Dependent variable: | |
|---|---|---|
| | ln(Earnings) | |
| | (1) | (2) |
| male | 0.280*** | 0.266*** |
| | (0.004) | (0.004) |
| age | | 0.110*** |
| | | (0.001) |
| I(age * age) | | -0.001*** |
| | | (0.00002) |
| Non_school | -0.190*** | -0.359*** |
| | (0.015) | (0.015) |
| No_HS_diploma | -0.233*** | -0.340*** |
| | (0.008) | (0.007) |
| No_college_degree | 0.220*** | 0.199*** |
| | (0.007) | (0.006) |
| Associate | 0.421*** | 0.316*** |
| | (0.009) | (0.008) |
| Bachelor | 0.760*** | 0.679*** |
| | (0.007) | (0.006) |
| Master | 1.070*** | 0.900*** |
| | (0.008) | (0.008) |
| Professional | 1.350*** | 1.148*** |
| | (0.015) | (0.015) |
| Doctor | 1.277*** | 1.064*** |
| | (0.016) | (0.016) |
| Married | | 0.199*** |
| | | (0.005) |
| Divorced | | 0.089*** |
| | | (0.008) |
| Separated | | -0.009 |
| | | (0.015) |
| Widowed | | 0.023 |
| | | (0.021) |
| Indian.Alaska | | -0.062*** |
| | | (0.015) |
| Asian | | 0.047*** |

| | | |
|---|---|---|
| | | (0.007) |
| Black | | -0.018* |
| | | (0.010) |
| White | | 0.129*** |
| | | (0.005) |
| Constant | 10.079*** | 7.456*** |
| | (0.006) | (0.026) |
| Observations | 144,748 | 144,748 |
| R2 | 0.237 | 0.353 |
| Adjusted R2 | 0.237 | 0.353 |
| Residual Std. Error | 0.823 (df = 144738) | 0.757 (df = 144728) |
| F Statistic | 4,992.987*** (df = 9; 144738) | 4,164.095*** (df = 19; 144728) |

Note: *p<0.1; **p<0.05; ***p<0.01
Numbers in () is robust standard error

## Table 3. Estimated ln(Earnings) Equations by Gender

| | Dependent variable: | |
|---|---|---|
| | ln(Earnings) | |
| | Male | Female |
| | (1) | (2) |
| age | 0.110*** | 0.109*** |
| | (0.002) | (0.002) |
| I(age * age) | -0.001*** | -0.001*** |
| | (0.00002) | (0.00002) |
| Non_school | -0.324*** | -0.433*** |
| | (0.018) | (0.024) |
| No_HS_diploma | -0.313*** | -0.405*** |
| | (0.009) | (0.012) |
| No_college_degree | 0.192*** | 0.201*** |
| | (0.008) | (0.009) |
| Associate | 0.288*** | 0.339*** |
| | (0.012) | (0.012) |
| Bachelor | 0.685*** | 0.661*** |
| | (0.008) | (0.009) |
| Master | 0.930*** | 0.865*** |
| | (0.011) | (0.011) |
| Professional | 1.137*** | 1.150*** |
| | (0.021) | (0.021) |
| Doctor | 1.054*** | 1.070*** |
| | (0.021) | (0.024) |
| Married | 0.302*** | 0.062*** |
| | (0.007) | (0.008) |
| Divorced | 0.122*** | 0.030*** |
| | (0.012) | (0.011) |
| Separated | 0.042* | -0.076*** |
| | (0.023) | (0.019) |
| Widowed | -0.008 | -0.028 |
| | (0.042) | (0.024) |
| Indian.Alaska | -0.077*** | -0.038* |
| | (0.020) | (0.022) |
| Asian | 0.003 | 0.101*** |
| | (0.009) | (0.010) |

| | | |
|---|---|---|
| Black | -0.062*** | 0.016 |
| | (0.014) | (0.014) |
| White | 0.139*** | 0.118*** |
| | (0.007) | (0.008) |
| Constant | 7.673*** | 7.540*** |
| | (0.035) | (0.039) |
| Observations | 81,301 | 63,447 |
| R2 | 0.369 | 0.323 |
| Adjusted R2 | 0.369 | 0.323 |
| Residual Std. Error | 0.762 (df = 81282) | 0.746 (df = 63428) |
| F Statistic | 2,640.275*** (df = 18; 81282) | 1,684.434*** (df = 18; 63428) |

Note:
*p<0.1; **p<0.05; ***p<0.01
Numbers in () is robust standard errors