Is College Worth the Money?
A Look on the Effects of Bachelor’s Degrees to the Unemployment Rate

By: Aatum Desai, Sameer Gir, Adam Nohl
ECON 3161, Spring 2017
Abstract

This paper aims to quantify the effect continuing with higher education after high school has on the unemployment rate. Fifty states were used as data points in this analysis to ascertain whether attending college greatly enhanced an individual's probability of being employed. Our hypothesis was that the greater the percentage of bachelor’s degrees to the population in a state, the lower that state’s unemployment rate would be. Other explanatory variables such as the percentage of people with a graduate degree, the increase in state GDP, the percentage of people with a high school degree, race distribution and median age were factored into the analysis to avoid data bias. Our data analysis shows a noteworthy correlation between the percentage of bachelor’s degree holders to the state’s unemployment rate - all other factors included.

I. Introduction

With the recent shift in political power from Democratic to Republican and the onslaught of mechanized jobs fueled by advances in technology, the unemployment rate has become a just cause for concern. As technology reduces the supply of basic jobs, the demand for jobs that require higher levels of education increases. Therefore, the question is raised in regards to how impactful completing a bachelor's degree in college is as opposed to leaving education after high school. Individuals are classified as unemployed if they do not have a job, have actively looked for work in the prior four weeks, and are currently available for work. The negative effects of unemployment reach far beyond the scope of those who are unemployed; often overlooked, failure to reach the natural rate of unemployment has costly effects on society and the country.

It is not difficult to fathom how being unemployed would affect an individual. Studies show that the average person would be in a crisis situation after only a few weeks without a paying job. Although the government has forms of assistance in place such as unemployment benefits and food stamps, this hardly amounts to the average person’s prior income. With significantly less income, people are forced to drastically reduce their consumption and potentially dig themselves into a deeper hole by having to borrow interest-burdened money or exhaust money set aside for retirement. Furthermore, the lack of savings can also hinder the progress of future generation by reducing the amount of available opportunities.

The rise of the unemployment rate is often coupled with an increase in blame towards immigrants for taking jobs that could potentially be held by citizens. This behavior not only encourages immigration bans but also leads to scenarios of trade restrictions and limitations. These affect society by decreasing diversity and harming local businesses that rely on suppliers or buyers overseas. Society is also affected when individuals may be more inclined to commit crimes in attempt to bridge the new gap in income.
The country is burdened by unemployment when states are required to support costly unemployment benefits. When taxes don’t fully cover these benefits, the government has a choice of borrowing money or redistributing their funding for other programs into the unemployment fund. Because unemployed individuals have less spending power, they buy less products which can actually harm the local economy more, perhaps even causing companies to lay off more workers. When these workers are laid off, the companies produce less, which hurts GDP.

Because the unemployment rate has such a significant impact on individuals and the country, it is carefully regulated by economic policies. Along with the effect of an extra four years of higher education has on the state unemployment rate, we believe that a state’s economic welfare, the distribution of the population’s race, median age and the percentage of the population with a high school degree and graduate degree are also significant factors tied to the unemployment rate.

II. Literature Review

In a March 2011 paper, researchers W. Craig Riddell (University of British Columbia) and Xueda Song (York University) investigate how an individual’s level of education affects their transition process between unemployment and re-employment. Most especially emphasized is the correlation between education and reemployment when unemployed- does a more educated person have greater ability to regroup and find employment? Riddell and Song (2011) begin by discussing the labor market and its constant overturn to make a case for the relevance of their studies. They cite that 10% of jobs are lost every year, but roughly a different 10% is also created (Davis and Haltiwanger, 1999). Since the world is continuously making rapid technological advancements that shift the size and disposition of the labor market, the threat of unemployment and the necessary development of soft skills are realities for employees in certain industries. The researchers state that based on past papers of a similar topic, they can hypothesize that greater levels of education lead to a greater incidence of reemployment after losing a job. However, their study differs from those done previously in that they are looking to eliminate other confounding variables, such as income, social networks, and natural intelligence. To address these concerns, they use data from a population that was put through compulsory schooling in the Vietnam War era- the 1980 census and the 1980-2005 Current Population Survey. Since these studies are mainly longitudinal, it gives a fine glimpse into the lives of those in the study and makes it easy to exclude individuals whose job prospects may be affected by extraneous factors. Additionally, they use instrumental variables (using compulsory schooling laws and child labor laws as instruments) instead of OLS regression, which additionally helps eliminate confounding variables. Per the estimates for the CPS
data, graduating from high school increased reemployment prospects by about 40 percent, and each additional year of schooling increases the probability by another 4.7 percent. However, they found mixed results for the relationship between level of education and chance of unemployment. There was no causal relationship at the secondary schooling level, but they did find a negative relationship at the postsecondary level.

In a second paper, Asoni, Andrea and Sanandaji (2016) use a multipronged identification strategy to determine whether college education is a good indicator to business and education survival. In it they state that a college education is indeed more beneficial to someone who is salaried rather than someone who is self-employed. To mitigate the endogeneity between education and ability, distance from college at the time of graduation and the local unemployment rate were put into a selection equation for college. With their data, it was determined that a higher education helps employees’ chances of employment. It was deemed that college may teach skills more valuable in a corporate environment than in an entrepreneurial one, and this conclusion was almost the same for women.

In another study, Chen, Guo, and Yu (2016) examine the effect of college on the self-employed. The study tries to determine the validity of a college education by examining how it affects entrepreneurship. Through multiple data regressions with self-employment as one and annual earnings as the other, it was seen that college education is positively correlated with being self-employed. Age, age squared, status of marriage, number of children, and state unemployment rates were kept as control variables, and having a college education improved the chance of being self-employed by 63.42%. The study also looks at two ethnic groups, Blacks and Hispanics. Again, it is seen within both groups that the relationship between a college education and self-employment is positive. When performing the regression in these cases, the two groups were considered as binomial variables where they equaled 1 when they belonged to that certain ethnic group and 0 otherwise. Typically Blacks and Hispanics have higher unemployment rates than other ethnic groups, but a college education severely lowers that rate as there is more chance of self-employment.

In Harmon’s 2003 paper, researchers focus on the individual rate of return achieved with investments in education. We can liken the rate of return as the chance of being employed. We find that for some countries there is a variation in returns between genders, the returns to women are significantly higher than the returns to men. However, this data may have been biased as it intrinsically depended on the participation rates of both genders. A smaller proportion of women participate in the surveys compared to men which may skew the results. They found a correlation between the participation rate and the gap in male and female returns. The higher the participation rate, the lower the differences. It was also seen that there was a positive relationship between returns and age and experience. A positive correlation
is also found between years of schooling including college and the returns achieved. Discrepancies between those at the upper end of the wage scale compared to those at the bottom were also discovered. Those at the top had a higher return than those lower. One of the reasons attributed to this may be the fact that there is also a positive correlation between education and ability. It concludes by mentioning that not only a degree will guarantee a graduate job but also other skills that may help in the job environment. The economy may even reach a point where there is actually overeducation in which case the benefit of education may be lowered as there is a saturation of high-skilled jobs and not enough low-skilled people available for jobs of like nature.

Our work will contribute to this literature and more like it by examining the more direct effect of a college education on the complete subject of employment, instead of its effect on specific parts of employment. This study will, like the above papers, be determining the value of a college education, but with simpler variables and controls so as to look at the results more completely and concretely. It’s unique because it will take into account more variables that will allow a broader perspective on the correlation.

III. Data

To find a correlation between completing a bachelor's degree and a state's unemployment rate, at the most rudimentary level, the percent of people with bachelor's degrees was regressed on the states. Along with these variables, the percentage of people with a graduate degree, the increase in state GDP, the percentage of people with a high school degree, race (binary - white or non-white) and median age were used to provide further insight. The 50 states are the data points used in the experiment (Washington DC is a significant outlier in terms of education and income and was consequently omitted.).

Simple Linear Regression

1. Unemployment rate

To conduct our study, unemployment rates by state were taken for the year 2015 as our dependent variable. This is defined as the percentage of people unemployed and looking for work to those who are in the labor force and employed. This data was taken from the Bureau of Labor Statistics. It is this variable that is the focus of our hypothesis. The regression equations are based on the primary independent variable’s correlation with this variable.

2. Percentage of People with a Bachelor's Degree at their Highest Level

For our simple regression model, the primary independent variable we have taken and the one around which we based the hypothesis for this paper is the percentage of people who have obtained a bachelor’s degree at their highest levels. This is the percentage of people in the state who have obtained a
bachelor’s degree and not continued further with their education. This was decided as the first independent variable because preliminary readings had shown that this variable seemed to have the strongest correlation with the unemployment rate. This study will determine whether this is true or not. This data was taken from the American Community Survey, data which is released by the United States Census bureau every year.

Multiple Linear Regression

3. Percentage of People with Graduate and Highest Degrees at their Highest Level

More independent variables were added to make the study more accurate in a multiple regression model. The percentage of people with a graduate degree was added as well as the percentage of people with a high school degree because we wanted to differentiate the two massive groups of people within the state. We wanted to distinguish these two categories from the those who had completed their highest level at only completed their bachelor’s degree. By distinguishing these variables, a more direct correlation and possible a causation will be seen between the primary independent variable and the dependent variable. Both datasets were taken from the American Community Survey for 2015.

4. Median Age

Median age was taken as another independent variable, as in the preliminary readings as well as through logical conclusion it was deemed that there might be a correlation between the median age of people in a state with the unemployment rate. It was not clear as to what effect the median age may have on the unemployment rate, but it through discussion it was decided that there was at least some correlation between these two variables. Data for ages was taken from Statista, a database company that compiles statistics from more than 18000 sources.

5. Race

Race in each state was also taken as an independent variable. This variable was a bit more complex as we had to consider it as a binary variable to use in our analysis and in our results. White people were considered as ones while the non-white population was considered as zero. This variable was considered because it was concluded that race may easily play a role in determining the employment status of someone, considering racial tensions in the United States. Data for this was taken from the Kaiser Family Foundation, a non-profit organization based in America.

6. Increase in State GDP

The last independent variable that was considered was the increase in state GDP. This was taken as the percentage of growth achieved by each state in 2016. This was taken as an independent variable as from previous studies, as it was suggested that the growth of the economy can have a substantial effect on
the unemployment rate. This would surely cause bias if not included as an independent variable in the multiple regression model. This data was taken from the Bureau of Economic Analysis for 2016.

From the summary statistics it can be seen that data was only taken from 50 states. Data from the District of Columbia was not taken because its inclusion caused the data to skew significantly. From the data, it is also seen that the proportion of people with a high school degree is highest while the proportion of people with a graduate degree is lowest in each state, an obvious conclusion given the cost of school.

Table 1 - Summary Statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment Rate</td>
<td>50</td>
<td>4.496</td>
<td>1.011961</td>
<td>2.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Bachelor’s Degree</td>
<td>50</td>
<td>18.278</td>
<td>2.769682</td>
<td>11.7</td>
<td>24.1</td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>50</td>
<td>10.736</td>
<td>2.600531</td>
<td>7.4</td>
<td>17.7</td>
</tr>
<tr>
<td>High school Degree</td>
<td>50</td>
<td>29.018</td>
<td>3.956534</td>
<td>20.7</td>
<td>40.7</td>
</tr>
<tr>
<td>GDP Growth Rate</td>
<td>50</td>
<td>3.402</td>
<td>1.438464</td>
<td>-0.1</td>
<td>7.1</td>
</tr>
<tr>
<td>Median Age</td>
<td>50</td>
<td>38.166</td>
<td>2.445454</td>
<td>30.6</td>
<td>44.6</td>
</tr>
<tr>
<td>White Percentage</td>
<td>50</td>
<td>0.6906</td>
<td>0.1602983</td>
<td>0.19</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Before undertaking an analysis of our results, a discussion of elements we changed from our initial model is needed. In terms of data used, we made two small changes. First, we omitted the District of Columbia from our analysis. Since the District of Columbia houses our federal government, it has a low population and a high amount of extremely educated, high-earning legislators, making it a high outlier in the realms of wealth and education. Secondly, in our initial model, we used a state’s GDP per capita as a measure of economic welfare; however, it had an infinitesimal coefficient (-0.00001), suggesting that it had little bearing on the model’s accuracy, so we substituted it for GDP growth instead. We believed that GDP growth was a better indicator of recent economic trends in specific states. In terms of regression variables, we made one notable change- we switched out Graduate Degree for High School
degree in the intermediate regression. Since we believed that the difference between a high school degree and a bachelor’s degree is far greater in terms of employment rate than a bachelor’s degree and graduate degree, we decided to change the variables to illustrate that significance.

Gauss-Markov Assumptions

The first Gauss-Markov assumption states that the model should be linear in parameters. This assumption justification is shown in the results section. The second assumption pertains to random sampling. Because the data was obtained from the American Community Survey and from the Bureau of Labor Statistics, who conduct annual surveys of randomly selected members of the population, it is safe to assume that the second assumption is met. The third Gauss-Markov assumption is the assumption of no perfect collinearity. From Table 3 in the results section, it is seen that the independent variables are not very strongly correlated to one another, therefore satisfying the assumption. As long as no two variables are perfectly collinear, this assumption will be met. The fourth assumption has to do with zero conditional mean; the error $u$ has an expected value of zero given any values of the independent variables. The last assumption is heteroskedasticity which also concerns $u$. As seen in Figures 1-3, the residuals show variances that vary randomly in each model. Also there is no discernable pattern in any of the figures. This indicates the assumption is satisfied.
IV. Results

Simple Regression

For simple regression analysis, the impact of higher education on the state’s unemployment rate was found using Equation 1,

\[ \text{Unemployment} = \beta_0 + \beta_1(\text{Percentage W/Bachelor's Degree}) + u \]

where \( \text{Unemployment} \) is the state’s unemployment rate, and \( \beta_1 \) represents the proportional change in \( \text{Unemployment} \) to any change of Percentage W/Bachelor’s Degree.

The regression was performed with STATA output being:

\[
\begin{array}{llllll}
\text{Source} & \text{SS} & \text{df} & \text{MS} & \text{Number of obs} = 50 \\
Model & 14.0016954 & 1 & 14.0016954 & F(1, 48) = 16.46 \\
Residual & 40.8311044 & 48 & .850648089 & \text{Prob > F} = 0.0002 \\
Total & 54.8328058 & 49 & 1.11903681 & R-squared = 0.2399 \\
\end{array}
\]

\[
\begin{array}{llllll}
\text{unempls} & \text{Coeff.} & \text{Sd. Err.} & t & \text{P>|t|} & [95\% \text{ Conf. Interval}] \\
\_cons & 8.539696 & 8.792408 & 9.71 & 0.000 & 6.771863 - 10.30753 \\
bdegree & -1.330023 & 0.475715 & -4.66 & 0.000 & -2.286512 - 0.973553 \\
\end{array}
\]

Based on this data, our simple linear regression model is:

\[ \text{Unemployment} = 8.5397 - 0.19300(\text{Bachelor's Degree}) + u \]
A graph of the least-squares regression line plotted over the data points is below.

**Figure 5: Simple Regression Model plot**

**Intermediate Regression**

For intermediate regression analysis, the impact of collegiate education, high school education, and GDP growth per state on the state’s unemployment rate was found using Equation 2,

\[
Unemployment = \beta_0 + \beta_1(\text{Percentage W/Bachelor's Degree}) + \\
\beta_2(\text{Percentage W/High School Degree}) + \\
\beta_3(\text{Increase In State GDP}) + u
\]

where *Unemployment* is the state’s unemployment rate, \( \beta_1 \) represents the proportional change in *Unemployment* due to any change of *Percentage W/Bachelor's Degree*, \( \beta_2 \) represents the proportional change in *Unemployment* due to any change of *Percentage W/High School Degree*, and \( \beta_3 \) represents the proportional change in *Unemployment* due to any change of *Increase in State GDP*. These variables were selected due to their hypothesized relevance - we simply selected the variables we believed would have the highest relevancy in terms of correlation. Additionally, none of the variables exhibited any perfect collinearity when regressed on each other.

The regression was performed with STATA output being:
Figure 6: Intermediate Regression Model Statistics

Based on this data, our intermediate linear regression model is:

\[ \text{Unemployment} = 13.5009 - 0.2668(\text{Bachelor's Degree}) - 0.1077(\text{High School Degree}) - 0.1506(\text{GDP Growth}) + u \]

Multiple Regression

For multiple regression analysis, we made use of every variable that we gathered data on and factored them into Equation 3,

\[
\text{Unemployment} = \beta_0 + \beta_1(\text{Percentage W/Bachelor's Degree}) + \\
\beta_2(\text{Percentage W/High School Degree}) + \\
\beta_3(\text{Increase In State GDP}) + \\
\beta_4(\text{Percentage W/Graduate Degree}) + \beta_5(\text{Median Age}) + \\
\beta_6(\text{Race}) + u
\]

where Unemployment is the state’s unemployment rate and \( \beta_1 \) through \( \beta_6 \) represent any proportional change on Unemployment caused by Percentage W/Bachelor’s Degree, Percentage W/High School Degree, Increase in State GDP, Percentage W/Graduate Degree, Median Age, and Race. When regressed on each other, none of these variables displayed perfect or near-perfect collinearity.

The regression was performed with STATA output being
Based on this data, our multiple regression equation is

\[
\text{Unemployment} = 10.5479 - 0.3777(\text{Bachelor’s Degree}) - 0.0983(\text{High School Degree}) - 0.0464(\text{GDP Growth}) + 0.1617(\text{Graduate Degree}) + 0.1002(\text{Median Age}) - 1.7135(\text{Race}) + u
\]

Table 2 - Statistical Inference

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>SIMPLE REG</th>
<th>INTERMEDIATE REG</th>
<th>MULTIPLE REG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor’s degree</td>
<td>Coef = -0.1930</td>
<td>Coef = -0.2658</td>
<td>Coef = -0.3777</td>
</tr>
<tr>
<td></td>
<td>T = -4.06***</td>
<td>T = -4.34***</td>
<td>T = -5.60***</td>
</tr>
<tr>
<td></td>
<td>SE = 0.0476</td>
<td>SE = 0.0613</td>
<td>SE = 0.0675</td>
</tr>
<tr>
<td></td>
<td>CI = (-.2887, -.0974)</td>
<td>CI = (-.3891, -.1425)</td>
<td>CI = (-.5138, -.2417)</td>
</tr>
<tr>
<td>High School degree</td>
<td>Coef = -.1077</td>
<td>Coef = -0.0983</td>
<td>Coef = -0.0983</td>
</tr>
<tr>
<td></td>
<td>T = -2.64**</td>
<td>T = -2.06**</td>
<td>T = -2.06**</td>
</tr>
<tr>
<td></td>
<td>SE = .0408</td>
<td>SE = .0478</td>
<td>SE = .0478</td>
</tr>
<tr>
<td></td>
<td>CI = (.1899, -.0256)</td>
<td>CI = (-.1947, -.0019)</td>
<td>CI = (.1899, -.0256)</td>
</tr>
<tr>
<td>GDP growth</td>
<td>Coef = -.1506</td>
<td>Coef = -0.0464</td>
<td>Coef = -0.0464</td>
</tr>
<tr>
<td></td>
<td>T = -1.66</td>
<td>T = -0.59</td>
<td>T = -0.59</td>
</tr>
<tr>
<td></td>
<td>SE = .0908</td>
<td>SE = .0792</td>
<td>SE = .0792</td>
</tr>
<tr>
<td></td>
<td>CI= (-.3335, .0322)</td>
<td>CI= (-.2062, .1134)</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>Grad degree</td>
<td>Coef = 0.1617</td>
<td>Coef = 0.1617</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T = 2.51**</td>
<td>T = 2.51**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE = 0.0645</td>
<td>SE = 0.0645</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI= (.0316, .2919)</td>
<td>CI= (.0316, .2919)</td>
<td></td>
</tr>
<tr>
<td>Median age</td>
<td>Coef = .1002</td>
<td>Coef = .1002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T = 1.76*</td>
<td>T = 1.76*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE = .0569</td>
<td>SE = .0569</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI= (-.0145, .2150)</td>
<td>CI= (-.0145, .2150)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>Coef = -1.7135</td>
<td>Coef = -1.7135</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T = -2.17**</td>
<td>T = -2.17**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE = .7892</td>
<td>SE = .7892</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI= (-3.3050, -1.220)</td>
<td>CI= (-3.3050, -1.220)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>Coef = 8.5397</td>
<td>Coef = 13.5086</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T = 9.71***</td>
<td>T = 6.59***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE = 0.8792</td>
<td>SE = 2.0512</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI= (6.7719, 10.3075)</td>
<td>CI= (9.3800, 17.6380)</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.2554</td>
<td>0.3906</td>
<td></td>
</tr>
</tbody>
</table>

Notes- all confidence intervals are 95%. One asterisk denotes that the statistic is significant at 10%, two denotes 5%, and three denotes 1%.

**Table 3 - Multicollinearity**

<table>
<thead>
<tr>
<th></th>
<th>Bachelor's degree</th>
<th>High School Degree</th>
<th>GDP Growth</th>
<th>Grad Degree</th>
<th>Median Age</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor’s</td>
<td>X</td>
<td>.4300</td>
<td>.1295</td>
<td>.4797</td>
<td>.0003</td>
<td>.0005</td>
</tr>
<tr>
<td>degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Degree</td>
<td></td>
<td>.0395</td>
<td>.2416</td>
<td>.1916</td>
<td>.1646</td>
<td></td>
</tr>
<tr>
<td>GDP Growth</td>
<td></td>
<td></td>
<td>.0054</td>
<td>.0005</td>
<td>.0598</td>
<td></td>
</tr>
<tr>
<td>Grad Degree</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>.0609</td>
<td>.0231</td>
</tr>
</tbody>
</table>
Initially, it appears that there is some multicollinearity between the bachelor’s, graduate, and high school degrees. However, this multicollinearity can be mitigated for 2 reasons: looking at the summary table, all of these variables have low standard errors, and all are significant at at least a 5% level in every study, with most instances even being significant at the 1% level. Even though some of the R-squared values are somewhat high, due to these factors, we are able to proceed with our analysis.

Analysis

Overall, our initial simple regression model had a low-to-moderate correlation (-0.2554) by itself, which states that a college education does have a large effect on one’s employment. Even though many jobs do not require a college degree, this regression confirms our initial hypothesis that a more well-educated population leads to a greater level of statewide employment. In the multiple regression, the two coefficients of HS degree and GDP growth certainly had a significant effect on the regression line, but the bachelor’s degree coefficient still had a higher absolute value than the two other coefficients added together, illustrating that the percentage of citizens with a bachelor’s degree is still a better predictor of unemployment than a highschool degree and GDP growth.

In the multiple regression analysis, race has a very sizable coefficient, but it is incomparable to the other variables in terms of effect since it is a binary variable and the other five variables are expressed as percentages. Omitting race, bachelor’s degree has far and away the highest absolute value of the coefficients (around 0.38, next highest value is around 0.16), and is also significant at the 1% level. This shows that out of the five non-binary variables we looked at, bachelor’s degree is the best predictor of a state’s unemployment rate. The variable with the next highest absolute value is the percentage of citizens with a graduate degree, which only confirms further our general hypothesis that a state’s focus on higher education directly relates to the percentage of its citizens that are unemployed.

The main dissension in this model came in the switching of graduate degree to high school degree in the intermediate regression, since the final regression showed that graduate degree interestingly had a greater influence on unemployment than high school degree. In hindsight, the intermediate regression would likely include bachelor’s degree, graduate degree, and race (when regressed individually, race had nearly a 0.2 correlation with unemployment rate).
Robustness Test - F-Test

Since the GDP growth rate and the percentage of people with a high school degree had the smallest coefficients in the multiple regression model, they were taken as the two variables in the restricted model of the robustness test to determine if they were indeed more useful to the model joint than they were individually. Their F-test value was 8.707, while their critical value was 2.589. This displays that these two variables are jointly significant and important to the multiple regression model.

V. Conclusion

In each model, it is seen that there is a negative relationship between the percentage of people who have bachelor’s degrees and the unemployment. This further corroborates the hypothesis posed at the start of this paper. As more explanatory variables were added, it was observed that the absolute value of the t-value for the bachelor’s degree as well as that of the coefficient’s increased. These results display that the primary explanatory variable is significant at the 1%, 5%, and the 10% levels. This further points to the multiple regression model as being the best representation of the hypothesis. It is observed that the percentage of people with a graduate degree is positively correlated. This may occur due to the fact, as mentioned in Harmon’s paper (2003), of overeducation occurring in these various states in that year.

While GDP growth rate and high school degree percentage may not have had the biggest correlations with the unemployment rate individually, using the robustness test it was discovered that these variables were indeed significant when considered together.

In this experiment we set out to determine how important it was for one to go the extra mile and complete a bachelor’s degree. To avoid biased results, we teased out other explanatory variables from the error term. These variables included the percentage of people with a graduate’s degree, the percentage of people with a high school degree, each state’s GDP growth, the distribution of race within each state, and the median age of the population in each state. The final multiple regression model yielded a correlation coefficient of -0.38 when comparing the percentage of the population with a bachelor’s degree to the state’s unemployment rate. With these kind of results, it is obvious that parents and politicians should push young adults to attend college and receive higher education, thereby reducing the unemployment rate.

References


