

Overeducation: The Effects of the Great Recession on the Labor Market

Alisha Gurnani

Columbia University

Graduate School of Arts & Sciences

GR 5999: Master's Thesis

Advisor: Dr. Elena Krumova

Advisor: Dr. Yao Lu

10 May 2021

Abstract:

This paper presents an analysis of overeducation rates in the United States prior to, during, and following the Great Recession of 2007 to 2009, in order to examine if there is an association between economic recessions and rising overeducation rates. Real Gross Domestic Product was used as a proxy variable to measure recession effects and education level data from the Census Bureau and the Bureau of Labor Statistics was used to measure the rate of Overeducation in the United States. The analysis was conducted using logistic regression models and concluded the possibility that the effects of a recession may increase the odds of an individual being overeducated. An additional sub-analysis was conducted to demonstrate the differences in overeducation rates by various subsets of groups in the United States, specifically across regions and age groups and found that regional differences in overeducation rates are fairly small and younger individuals have higher rates of overeducation than older individuals.

Introduction:

Educational training is considered to be an important aspect in providing individuals in the labor force with the training necessary to succeed in their chosen field. In the U.S., educational attainment at every level has been increasing steadily for the past several years (US Census Bureau, 2020). This suggests that there are more educated individuals that enter the labor market each year, but to meet this increase in labor supply there also needs to be an increase in labor demand, which may not always occur. This can lead to many individuals accepting a job offer that they are overqualified for in terms of skills, training, experience, and/or education. Specifically, when an individual possesses a higher education than that required of the job, then he or she is considered to be overeducated.

Overeducation is a phenomenon that has been studied for several years and has continued being studied in the United States because of the potential negative effects overeducation may have on individuals and on labor productivity, such as decreased life satisfaction and decreased productivity. Many researchers have examined overeducation over long periods of time, between immigrants and non-immigrants, within different age groups, etc. However, overeducation through the period of a recession has not been studied as thoroughly.

During and after a recession the labor market dynamics change considerably. There are more constraints, since there are fewer job openings than before and a larger labor supply, leading to the possibility of individuals in the labor force accepting jobs they are overqualified for. This relationship between overqualification and recessions is what this paper will be examining to provide some further insight into the overeducation phenomenon; recession will be measured using Real Gross Domestic Product of the U.S. as a proxy variable and overeducation, will be constructed using education data.

The explicit research question this paper intends to explore is: how does overeducation in the labor market change after an economic recession? I hypothesize that overeducation would increase after an economic recession, due to several reasons ranging from increased competition in the job market to a decrease in individual motivation in actively job searching. However, my intention of this paper is not to find a cause for overeducation, but to see whether a rise in overeducation is correlated to economic recessions. While conducting the study, I will specifically look at the effects of the Great Recession, which occurred from 2007 to 2009, on overeducation, specifically studying the time period 2004-2012 to observe overeducation before, during, and after the recession. If overeducation and the proxy variable, GDP, exhibit a negative association either in during or after the recession, it would convey a positive relationship

between overeducation and recessions, suggesting that overeducation rises as a result of recession effects.

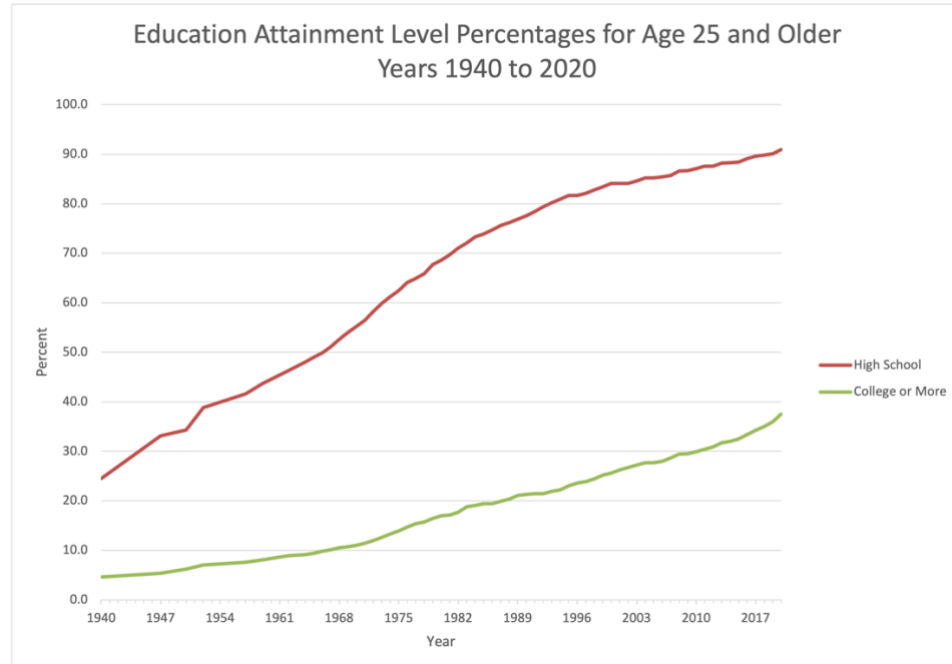
Additionally, I will further investigate a subset of groups to see if overeducation occurs more or less in certain subgroups. The two subsets I will be looking into are geographical regions in the United States and age groups. Specifically, I'd like to see if overeducation occurs more in rural regions of the U.S. because rural regions are more likely to have a lower demand for labor and if overeducation occurs more in younger individuals such as individuals aged 35-45 since past research indicates that "overeducation falls with increasing age and experience" (Hartog, 2000).

Literature Review:

Defining Overeducation:

Educational attainment in the United States has been steadily increasing since the 1940s, when the U.S. Census Bureau first released the question through their census questionnaire (US Census Bureau, 2012). Additionally, the percentage of individuals with a bachelor's degree or more in general and in the labor force have increased consistently from 1992-2016 as seen in the Figure 1 below (Brundage, 2017).

Figure 1.



Note. This chart shows the percentage of people aged 25 years and over who have completed either High School or College or more through the years of 1940 to 2020 (US Census Bureau, 2021).

While the education requirements for occupations may change through the years with the creation of new jobs and industries, it can be speculated that a percentage of the workforce are still overeducated. Additionally, many jobs have had an “upgrade in the skills needed to perform [tasks] adequately” creating a demand for higher-educated workers however “if the growth in supply of higher-educated workers outpaces the growth in demand, overeducation of the workforce is the likely result” (Groot & Maassen van den Brink, 2000).

While there are several forms of mismatched skills prevalent in the labor force such as undereducation, that refers to an individual having less education than required, or over skilled, which refers to possessing more skills than required regardless of how it was acquired, over education will be the focus of this paper (Capsada-Munsech, 2017). Overeducation refers specifically to the concept that the level of education an individual has is greater than that of

what is required for their occupation (McGuinness, 2006). In other words, overeducation can be defined as a mismatch of acquired skills versus required skills (Capsada-Munsech, 2017).

Past research has shown that the incidence of overeducation is particularly high in the United States, especially among immigrants (Ramos, 2013; Lu & Hou, 2020). For instance, a paper by Beckhusen et al. found that over 40% of immigrants with a bachelor's degree are overeducated, while their native counterparts have an overeducation percentage that is 10 to 20 points lower but overall, the rate of overeducation is high in the United States (2013).

Overeducation and Economic Recessions:

To examine whether there is a correlation between financial crises and overeducation, first, defining what a recession is and establishing what can be an appropriate proxy to measure a recession is necessary. Recessions are defined as a “significant decline in economic activity spreading across the economy, lasting more than a few months” and can be measured by “domestic production and income [measures], such as gross domestic product, gross domestic income, and industrial production” (Smith, 2009). For this paper, I will be using the gross domestic product (GDP) of the United States from the years 2004 to 2012 as a way of measuring condition of the economy before, during, and after the Great Recession of 2007-2009.

Recessions can have several negative effects. For example, a widely acknowledged consequence is that an economic recession can cause a rise in unemployment rates due to the changes in the labor market that occur (Sánchez-Sánchez & Fernández Puente, 2020; *The Recession of 2007–2009: BLS Spotlight on Statistics*, 2012). In the United States, the unemployment rates rose from 5% to 9.5% from December 2007 to June 2009 and peaked at 10% after the recession (*The Recession of 2007–2009: BLS Spotlight on Statistics*, 2012). Based on the knowledge that

changes in the labor supply and demand are related to changes in unemployment rates can logically lead to the hypothesis that changes in the labor market are related to changes in overeducation rates (Sánchez-Sánchez & Fernández Puente, 2020). This is because “overeducation is fundamentally a product of labor market dynamics” (Borgna et al., 2018). During an economic downturn, the demand for labor decreases whereas the supply of labor increases. This “lack of employment [opportunities] could force individuals to accept a job, regardless of the alignment with their levels of studies” (Sánchez-Sánchez & Fernández Puente, 2020). There are not many studies that look at the relationship between economic recessions and overeducation in the United States, however a few researchers have conducted similar studies to examine the effects of the Great Recession on overeducation in Europe.

A study done by Borgna et al. examined the differences in the rates of overeducation between individuals who switched employers after the recession and those who stayed at the same employer for a long period of time throughout Europe (Borgna et al., 2018). They studied how poorer economic conditions affected both of these types of workers, specifically focusing on prime age workers, which were defined as 35–55-year-olds (2018). Their results concluded that “poorer economic conditions are associated with higher overeducation risks, especially for those who switched employers after the outbreak of the crisis” (2018).

Sánchez-Sánchez & Fernández Puente conducted a study on the Spanish labor market dynamics and examined the overeducation rates from 2006-2013 to see if they increased (Sánchez-Sánchez & Fernández Puente, 2020). The study also examined if there was any persistence and inertia in the overeducation rates because of the possibility that “past overeducation [could affect] the likelihood of current overeducation” (2020). They utilized the European Union Survey of Income and Living conditions that provided data in panel form for

the years 2006-2009 and 2010-2013 but not the full time period being studied; to account for this they studied the two periods of panel data independently (2020). To calculate overeducation they used the objective method of evaluating one's education level which consists of comparing the level of education attained by the individual to the occupation education requirements. Through their analysis they found that "the percentage of overeducation increases from 11.91 to 13.95% from 2006 to 2009, although not continuously, and from 14.74 to 16.41% from 2010 to 2013" (2020). The study also concluded that there was inertia and persistence in Overeducation because they found that "being overeducated at the beginning of the period (2006) [increased] the probability of being overeducated [at the end of the period]" (2020).

Potential Impacts of Overeducation:

Overeducation in the labor market can have a negative impact on the workers as well as the overall economy, which is why it's important to continue studying the phenomenon and how it might relate to other variables, such as the business cycle, age, geographic location, etc. A potential consequence of overeducation in the labor market is that "at a macroeconomic level, national welfare [could be] lower than would be the case if the skills of all overeducated workers were fully utilized within the economy" (McGuinness, 2006). Additionally, McGuinness explains that there could be effects at the company level as well, such as lower overall productivity for the firm (McGuinness, 2006).

A review of the literature has shown that most of the negative effects occur at the individual level. Individuals who are overeducated are more likely to have a wage penalty compared to individuals who are correctly matched in their occupations (Meroni & Vera-Toscano, 2017). Similarly, overeducated workers receive a lower return on their investment,

since they are not compensated accordingly along with “the fact that a proportion of their educational investment is unproductive” (McGuinness, 2006).

Moreover, overeducated workers are more likely to be less satisfied with their jobs leading to lower productivity levels (Beckhusen et al., 2013). They are also more likely to be less satisfied with life. According to one study done by Artes et al., they used a self-reporting method to gather satisfaction levels of overeducated and under-educated individuals and found “over-educated people report slightly lower levels of life satisfaction compared to under-educated people” (2014).

For recent graduates, the effects of overeducation can be long-lasting. The younger generation has been increasingly encouraged to pursue higher education, with one paper stating that “in the US [...] the percentage of 25-30-year-olds with at least a bachelor’s degree increased by about 4% points, from 29.1% in 2000 to 34% in 2014” (Waldorf & Yun, 2015). And although, many graduates are able to find jobs that match their education level, many also become unemployed or decide to take jobs they are overeducated for out of fear of unemployment (Waldorf & Yun, 2015). The negative impacts of taking a job an individual is overeducated for early on includes “sending negative signals to future employers” and diminishing the potential job experience that the individual could have gained (2016). Furthermore, more research on the negative impacts of overeducation on recent graduates have found that there is a possibility for the trap hypothesis, which says that “accepting a position for which one is overeducated is always associated with less likelihood of ending up in a matched on later on” (Meroni & Vera-Toscano, 2017).

Measuring Overeducation:

Currently, there is no universally accepted method of measuring overeducation; most often the method utilized is dependent on what data is available to the researcher (Hartog, 2000). One paper discusses four of the different methods that are considered by researchers when calculating overeducation however none are preferred over the other (Groot & Maassen van den Brink, 2000). The four methods can be divided into two categories: subjective and objective (Groot & Maassen van den Brink, 2000). The two subjective methods consist of a self-report, containing one question each, that is made by the workers, where the workers are either asked whether they are over or under-educated or they are asked what the minimum education requirement is for individuals at their level of occupation (Groot & Maassen van den Brink, 2000). Although, the worker assessment report is beneficial in terms of gathering current information regarding education requirements, it can potentially contain workers' biases (Hartog, 2000). The two objective methods are either comparing the attained education level of an individual to the average education level of workers within the occupation or comparing the level of education attained by the individual to the required education level of the occupation (Groot & Maassen van den Brink, 2000). The objective methods also have their advantages and disadvantages. Since the goal of the education and occupation classifications used in the two latter methods is to be objective, through "clear definitions and detailed measurement instructions", there is a low chance for the data to be biased (Hartog, 2000). However, the education and occupation classifications are based on the "assumption that workers with the same job or occupation title are doing work of equal difficulty", which may not always be the case (Dolton & Vignoles, 2000). For this paper, I will be focusing on using the second objective method discussed above: comparing one's level of education to that of the education required by

the occupation. This is due to the availability of occupation and education classification data through federal sources and to ensure any bias effects are minimized.

Data Description:

To measure overeducation in the United States, I constructed the variable using two datasets that help determine what the requirements are for each occupation and what education level individuals have obtained. I used data from the American Community Survey (ACS) as well as data from the Bureau of Labor Statistics (BLS). The first dataset, the ACS data, is survey data that is collected by the U.S. Census Bureau, which is released as either a 1-year profile or a 5-year profile for the country. The ACS contains demographic data for the United States, such as Region, Age, Sex, Educational Attainment, Employment, Occupation, and more. For the purposes of this study, I used the ACS 1-year profiles from 2004-2012 through the IPUMS database. IPUMS provides the ACS. The IPUMS database curates and distributes ACS microdata from population censuses; this data consists of sample data of about 1% to 5% of the national population. I restricted the sample to include full observations and only those that have listed their employment status as “Employed” for the years being studied to make sure only the active labor market is captured in the analysis.

The second source of data is the 2010 BLS Employment Projections data. This data is based on a Standard Occupational Classification (SOC) system by the BLS that determines the requirements for each occupation. The dataset contains variables on occupations in the United States and the typical education degrees required for each job, as well as how much work experience may be needed. The BLS assigns the typical education required for each occupation by the following categories: doctoral or professional degree; master’s degree; bachelor’s degree;

associate's degree; postsecondary nondegree award; some college, no degree; high-school diploma or equivalent; or no formal education credential.

To examine overeducation before, during, and after the Great Recession, a new variable named Period was created to split the multiple years of data the three respective time periods. The years 2004 to 2006 is labelled as Period 1 for before the recession, the years 2007 to 2009 is labelled as Period 2 for during the recession, and the years 2010 to 2012 is labelled as Period 3 for after the recession.

To construct the overeducation variable, I used the two sources of data listed above and used the measurement method mentioned in the above literature review, the objective method that requires conducting comparisons. The ACS survey data were first preprocessed to account for any missing data and were subset to only include individuals that reported being employed in the current survey year. Since individuals may not answer all of the questions asked, only rows with complete data were used. Using the processed ACS data from IPUMS and the BLS standard occupational classification system, the level of education obtained by the individual was matched to the level of education required for the occupation by the SOC occupation ID. In the IPUMS data there are some occupation ID's that are missing the last few identification digits, so to make sure those were still accounted for I conducted the matching process with the ACS occupation ID using the crosswalk data for the BLS SOC and ACS. The overeducation variable is coded in a binary manner, where 1 equates to overeducation, and 0 equates to no overeducation. In the table below are the percent distributions of individuals in the labor market who are overeducated and individuals who are not (See Table 1).

Table 1: Percent of Overeducated and Not Overeducated by Year

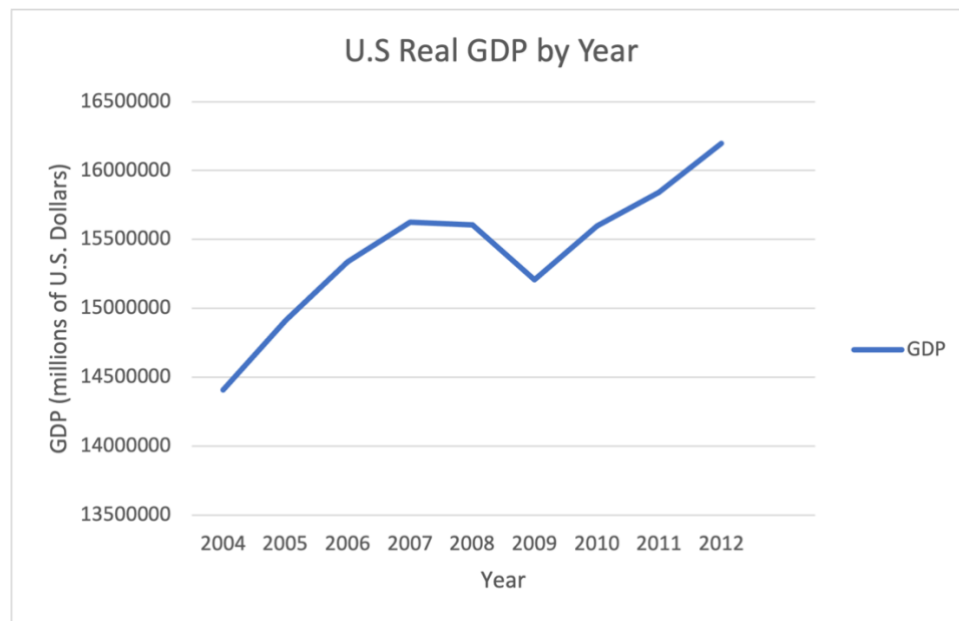
Year	Overeducated	Number	Percent
2004	0	182954	50.35
	1	180391	49.65
2005	0	440839	50.66
	1	429391	49.34
2006	0	457432	50.56
	1	447227	49.44
2007	0	456828	50.41
	1	449336	49.59
2008	0	453199	49.08
	1	470253	50.92
2009	0	426157	48.32
	1	455803	51.68
2010	0	442434	47.79
	1	483436	52.21
2011	0	435976	47.56
	1	480732	52.44
2012	0	431660	46.87
	1	489332	53.13

To measure the economic recession of 2007-2009, the independent variable, I used the Gross Domestic Product (GDP) of the U.S as proxy from the years 2004 to 2012. The dataset I used for this variable is from the Bureau of Economic Analysis (BEA). The BEA provides both real and nominal GDP as well as seasonally adjusted and non-seasonally adjusted GDP. For this study I used seasonally adjusted real GDP to account for any seasonal fluctuations through each year and any inflation that occurred over the time period I chose to study, to better reflect the economic patterns of the United States before, during, and after the recession. The GDP data is released quarterly, and the annual GDP data is calculated using a formula similar to a compound interest formula. The annual GDP data will be used for the analysis in this paper. Below are the

yearly GDP numbers provided by the BEA. The GDP decreases in 2009 indicative of the after-effects of the crash that occurred in 2007.

(See Figure 2)

Figure 2.



Note: This chart shows the seasonally adjusted annual Gross Domestic Product for the United States during the years 2004-2012 in millions of U.S. dollars (Bureau of Economic Analysis).

Methodology:

I conducted multiple logistic regression models to determine whether there was a significant correlation between overeducation and GDP. This specific type of regression was used because the dependent variable, Overeducation, is binary and would be best examined through a logistic regression.

The first model is a basic model which looks at the relationship between the dependent variable, Overeducation and the main independent proxy variable, GDP, without any control variables. This model provides an initial understanding of the relationship between GDP and

Overeducation. This basic regression was used as a baseline model for the rest of the research conducted.

The second logistic regression model included the Period variable as a control variable to understand the relationship between GDP and overeducation through the three time periods. This variable was included as a factor variable to examine the log odds of overeducation in each time period: before, during, and after the recession.

In the third logistic regression model I included a few other control variables in addition to the Period variable. Specifically, Age, Gender, and Education Attainment levels were incorporated to explore the relationship between GDP and overeducation while keeping other variables that may be correlated to overeducation constant. An additional fourth model was conducted without the Education Attainment level control variable, to explore how this might affect the magnitude and significance of the correlation coefficient on GDP.

Lastly, after producing results from the final logistic regression model, I plotted the results by age groups to explore how overeducation in the labor market differs by age. The age groups will start at minimum employment age and will be grouped similarly to the age buckets in the labor force data by the BLS as follows: 16-24, 25-34, 35-44, 45-54, 55-64, 65 and up. Additionally, I will plot the results by regions of the United States to explore how overeducation differs across the regions. The regions will be set based on the ACS regional definition as follows: Far West, Great Lakes, Mideast, New England, Plains, Rocky Mountain, Southeast, and Southwest.

Results:

Regression Results:

Table 2 below includes the base line logistic model (Model 1) results with the single independent variable, GDP, and the dependent variable, Overeducation. The results show that GDP ($9.02e^{-8}$) has a small but statistically significant positive correlation with Overeducation when the logistic regression is conducted with no control variables. Since GDP is measured in millions of dollars, the coefficient on GDP can be interpreted as such: for every one million U.S. dollar increase in GDP, the log odds of Overeducation increases by $9.02e^{-8}$. The positive relationship suggests that as GDP is rising, or as the economy is improving, the log odds of an individual being overeducated also rises.

Table 2: Logistic Regression Results with GDP
Model 1

	<i>Dependent variable:</i>
	Overeducation
Gross Domestic Product	0.0000000902*** (0.0000000017)
Constant	-1.3561480000*** (0.0260256200)
Observations	7,613,380
Log Likelihood	-5,274,100.0000000000
Akaike Inf. Crit.	10,548,205.0000000000
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3 below consists of the logistic regression results from Model 2 which includes Overeducation, GDP, and Period to account for the time periods of before, during, and after the recession. The correlation coefficient on GDP ($-1.06e^{-8}$) in Model 2 becomes negative after

controlling for time periods, meaning that as GDP decreases, the log odds of being overeducated rises. This relationship with the proxy variable insinuates that there may be a positive correlation between recessions and overeducation. Period 2 and 3 both result in a positive, statistically significant, correlation coefficient. In Period 2 the correlation coefficient is 0.057 indicating that the log odds of being overeducated in Period 2 increases by 0.057 relative to Period 1. In Period 3 the correlation coefficient is 0.136 indicating that the log odds of being overeducated in Period 2 increases by 0.136 relative to Period 1. These results seem to imply that the highest chance of an individual being overeducated occurs in Period 3, in other words the highest chance of being over educated occurs after a recession.

Table 3: Logistic Regression Results with GDP and Time Period
Model 2

	<i>Dependent variable:</i>
	Overeducation
Gross Domestic Product	-0.0000000106*** (0.0000000028)
Period 2	0.0566300000*** (0.0022694630)
Period 3	0.1357993000*** (0.0030597690)
Constant	0.1366483000*** (0.0423268100)
Observations	7,613,380
Log Likelihood	-5,273,041.0000000000
Akaike Inf. Crit.	10,546,091.0000000000
<i>Note:</i>	* p<0.1; ** p<0.05; *** p<0.01

Table 4 below contains the Model 3 logistic regression results, which includes the dependent variable, Overeducation, and several control variables, GDP, Period, individual

Education Attainment Level, Age, and Gender. The coefficient on GDP ($4.4e^9$) in Model 3 is now positive again like it was in Model 1 and is no longer statistically significant. This suggests that although there is a positive relationship between GDP and overeducation, GDP is not considerably associated with the rise in the log odds of overeducation. The coefficient on Period 2 is 0.034, meaning that during the Great Recession the log odds of being overeducated increased by 0.034 relative to the log odds in Period 1 and with all other variables being constant. This coefficient decreased in magnitude in Model 3 but is still statistically significant. The coefficient on Period 3 is 0.091, meaning that after the Great Recession the log odds of being overeducated increased by 0.091. This coefficient is also smaller in magnitude in Model 3 than in Model 2 but is still statistically significant. However, since GDP here is positive and not negative while the coefficients on Periods 2 and 3 are also positive, there is not enough evidence to assume that the recession effects correspond with the rising incidence of overeducation rather than another variable such as, Educational Attainment.

Education Attainment Level has the highest correlation coefficient in this model at a positive 0.512 and is statistically significant. This suggests that as an individual's education level increases their log odds of being overeducated increases by 0.512. In other words, as an individual's education level increases, they are more likely to be overeducated in the labor market.

The other control variables in this model, Age and Gender are also small but statistically significant implying that there is a relationship between Overeducation and these variables. Additionally, Age is a negative number, indicating that log odds for being overeducated increases if you are young.

Table 4: Logistic Regression Results with control variables
Model 3

	<i>Dependent variable:</i>
	Overeducation
Gross Domestic Product	0.0000000044 (0.0000000030)
Period 2	0.0335098900*** (0.0024511980)
Period 3	0.0907446000*** (0.0033048570)
Education Attainment Level	0.5121694000*** (0.0005234166)
Age	-0.0109372000*** (0.0000558640)
Gender: Female	0.0394478100*** (0.0015757950)
Constant	-1.3003690000*** (0.0457970000)
Observations	7,613,380
Log Likelihood	-4,702,754.0000000000
Akaike Inf. Crit.	9,405,522.0000000000
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

When removing Education Attainment Level from the model (as seen in Model 4 results below) the coefficient on GDP ($-1.00e^8$) is negative and statistically significant again. Which alludes to the notion that Education Attainment may have a more significant association with Overeducation than GDP does and so when controlling for it (as done in Model 3), GDP no longer has an important relationship with Overeducation. The coefficients on Periods 2 and 3 are consistent with the previous models, both are positive and show that during and after a recession the log odds of overeducation rises. The negative correlation between GDP and Overeducation in

Model 4 and the positive correlations on Periods 2 (0.058) and 3 (0.139) imply that there may be some association between recessions and the rise of overeducation rates.

Table 5: Logistic Regression Results with GDP, Time Period, Age, and Gender
Model 4

	<i>Dependent variable:</i>
	Overeducation
Gross Domestic Product	-0.0000000100*** (0.0000000028)
Period 2	0.0579680900*** (0.0022714750)
Period 3	0.1390531000*** (0.0030630230)
Age	-0.0040051090*** (0.0000515015)
Gender: Female	0.1111042000*** (0.0014595620)
Constant	0.2449380000*** (0.0424209300)
Observations	7,613,380
Log Likelihood	-5,266,930.0000000000
Akaike Inf. Crit.	10,533,873.0000000000
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Subgroup Evaluation Results:

As an addition to the above regression analysis, I chose to look at two subgroups in the United States and visualize the distribution of overeducation in those groups over time to provide a better understanding of the incidence of overeducation. The two subgroups I chose to further analyze were age groups, since there were many papers that discussed age as a significant variable in overeducation rates, as well as regions of the United States to see if there were any

geographical differences in the overeducation rates that reflected any differences in the labor markets across the country.

Figure 3 below displays the percent of overeducation per age group in the labor force through the years of 2004-2012. The percentages among all of the age groups are fairly similar however in the years after the financial crash of 2007, the percentage of individuals overeducated in the age groups 16 to 24 years and 25 to 35 years can be seen increasing. This graph suggests that younger individuals, ones who have recently graduated or are still new to the labor market, may be more likely to be overeducated. This poses the possibility that the rise in overeducation rates in younger individuals was indirectly affected by the Great Recession due to the shift in labor market dynamics as well as the possibility that employers were now requiring more years of experience for an occupation than before which may have disadvantaged the younger generation.

Figure 3.

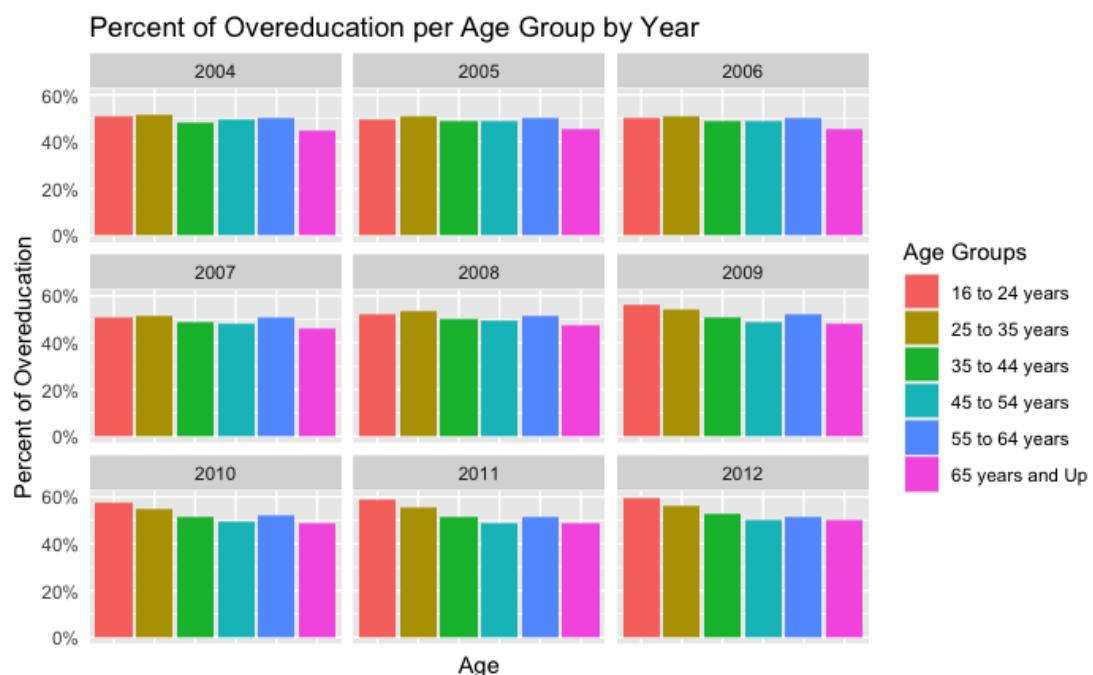
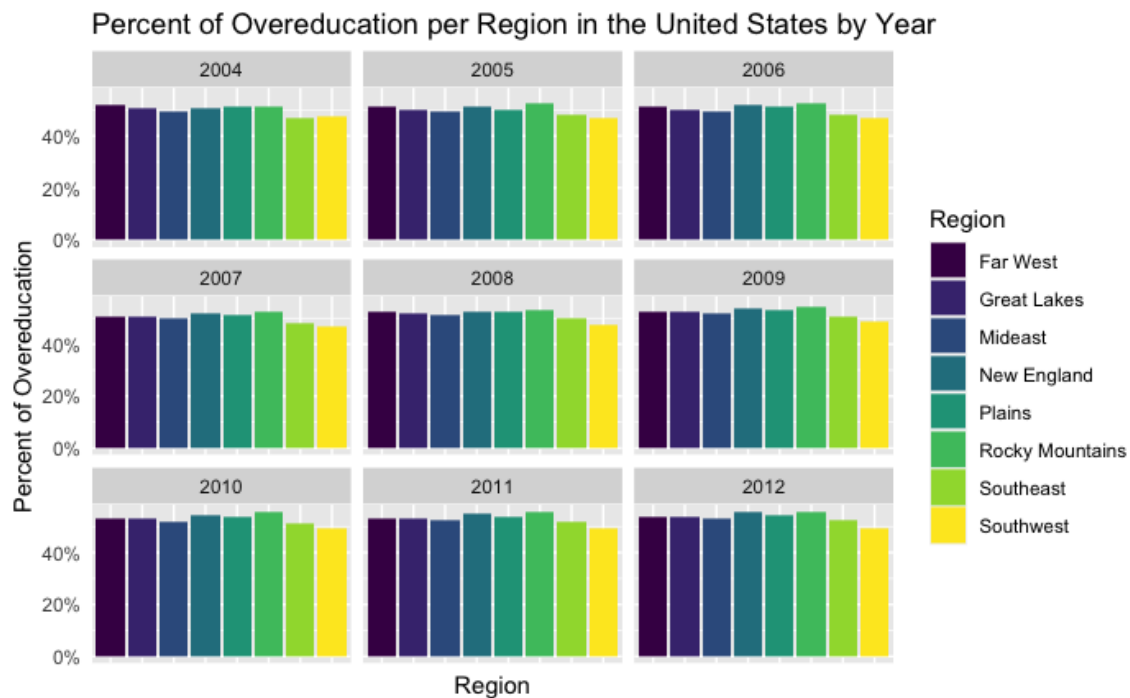


Figure 4 below displays the percent of overeducation in various regions throughout the United States over the years of 2004-2012. The regions used for this visualization are the ones defined by the Bureau of Economic Analysis. This graph as well shows fairly similar percent values for overeducation for all of the different regions per year, but it also seems to be fairly consistent before, during, and after the years of the Great Recession. The Southeast and Southwest regions consistently display lower percentages of Overeducation through the years, while the Far West, New England, Plains, and Rocky Mountain regions consistently display higher percentages of overeducation through the years. This may be due to the various occupations available throughout the United States all requiring different education levels, however there is not enough information to assume that as true.

Figure 4.



Discussion:

This paper examines the relationship between recessions and the prevalence of overeducation in the United States by analyzing the effects of the Great Recession of 2007-2009 on overeducation. Using a proxy variable, GDP, to measure the dip in the economic cycle and using a time periods variable as a means to distinguish the periods before, during, and after a recession, there was some evidence to conclude that recessions and overeducation have a positive relationship.

The results of the base model, Model 1, presents a general understanding of the relationship between GDP and Overeducation, but the results of Model 2, Model 3, and Model 4 provide a deeper understanding of the relationship between GDP and Overeducation when controlling for the time periods of a recession. The first and third model are consistent in their results in terms of the relationship between GDP and overeducation, both display a positive relationship, which suggests that as GDP increases the log odds of being overeducated increases. The second and fourth model are also consistent in their results, both showing a negative statistically significant relationship between GDP and overeducation. Furthermore, in the third model the coefficient on GDP is no longer significant after more control variables were included, specifically Education Attainment Level. This means that although there is a small positive correlation after the controls were added, either GDP is not significant enough to affect the incidence of overeducation in the United States or it illustrates that it may not be the best proxy variable to measure recessions and a more comprehensive variable, such as a well combined indicator consisting of GDP, Unemployment Rate, and the Consumer Price Index, might be a better approach to measure stages in an economic cycle.

The second, third, and fourth models consistently show a positive relationship between Period 2 and Overeducation, and Period 3 and Overeducation, both implying that during and after a recession the log odds of being overeducated rises, even after controlling for other variables that could affect the incidence of overeducation in the United States. This provides some evidence that the effects a recession has on the economy are correlated to the rise in overeducated individuals in the United States. However, this also may suggest that overeducation might be rising over these periods not as a parallel to recession effects but as a result of steady increases in individual educational attainment levels.

Nonetheless, it is best to remember that the magnitudes of all of these correlation coefficients are small, so although these findings give a little bit of understanding into the overeducation phenomenon, more research on this topic should be conducted for a more in-depth comprehension.

The insights from this paper add to existing knowledge of how recessions can potentially affect the economy and initiates the conversation on possible solutions of overeducation not just after a recession but more broadly as well. The insights also provide a basis for further questions such as what better variables could be used to indicate a recession, or what the relationship between recent graduates and overeducation is during and after a recession, or whether persistence and inertia play a role in the incidence of overeducation in the United States and if so, how big of a role would it play during the time period of recession, similar to the study Sánchez-Sánchez, N., & Adolfo Cosme Fernández Puente conducted in Europe.

Furthermore, additional research into the incidence of overeducation in different subsets of groups, such as gender, industry, race, through the period of a recession could be conducted to better understand the nuances of the overeducation phenomenon. Visualizing multiple layers of

variables such as displaying overeducation rates by race across regions in the United States over time, could provide more insight on how multiple aspects of an individual's demographic could potentially affect their chances of being overeducated.

Limitations:

There are several limitations to this study. First, although federal open data was used to calculate the overeducation variable, there are several methods of measuring overeducation and few methods of verifying if the measurement is accurate, which can lead to discrepancies in the results when compared to other overeducation analyses. Another limitation of the study is that only GDP is being used to measure an economic recession, however there are several indicators that show a recession is occurring or has occurred. For example, two other recession indicators are real income and employment rates; real income declines during a recession and unemployment rates rise. All of these measures could be used to provide a more holistic picture of the economic recession of 2007-2009. There is also a limitation of the data; there is a possibility that there are underrepresented and overrepresented groups of the population in the ACS 1-year data which could be reflected in the analysis results, especially when conducting the sub-group analysis. Finally, to conduct the matching process I had to utilize the 2010 BLS SOC classification system even for the years prior because there was not a comprehensive classification system before 2010 by the BLS that contained only education requirements for each occupation. The prior classification systems combined both training and education requirements into one categorical variable.

Conclusion:

The models conducted show a small negative relationship between GDP and overeducation in the United States during the Great Recession, indicating for the possibility that the effects of a recession may affect overeducation rates. Studying the effects of a recession on overeducation provides more insight into the effects a recession has on employed individuals and thus how this may be unfavorable for companies and the country's economy. Other independent variables are also significantly correlated with a rise in Overeducation, such as Educational Attainment, which suggests that overeducation is an important phenomenon to study especially since education attainment has been rising consistently in the United States in the last 20 years (US Census Bureau, 2020). These results and understanding of the relationships provide for a better base when devising public policies regarding education and the economy. Overall, this paper only provided a glimpse into the relationship between recessions and overeducation, and more studies would need to be conducted to get a more thorough understanding of the relationship.

References:

- Artés, J., Salinas-Jiménez, M. M., & Salinas-Jiménez, J. (2013). Small Fish in a Big Pond or Big Fish in a Small Pond? The Effects of Educational Mismatch on Subjective Wellbeing. *Social Indicators Research*, 119(2), 771–789. <https://doi.org/10.1007/s11205-013-0526-1>
- Beckhusen, J., Florax, R. J. G. M., Poot, J., & Waldorf, B. S. (2013). ATTRACTING GLOBAL TALENT AND THEN WHAT? OVEREDUCATED IMMIGRANTS IN THE UNITED STATES. *Journal of Regional Science*, 53(5), 834–854.
<https://doi.org/10.1111/jors.12030>
- Borgna, C., Solga, H., & Protsch, P. (2018). Overeducation, Labour Market Dynamics, and Economic Downturn in Europe. *European Sociological Review*, 35(1), 116–132.
<https://doi.org/10.1093/esr/jcy046>
- Brundage, V. (2017, August 17). *Profile of the labor force by educational attainment : Spotlight on Statistics: U.S. Bureau of Labor Statistics*. Bureau of Labor Statistics.
<https://www.bls.gov/spotlight/2017/educational-attainment-of-the-labor-force/home.htm#:~:text=Labor%20force%20participation%20rises%20with%20education&text=In%202016%2C%20the%20labor%20force,percent%20and%2075.1%20percent%2C%20respectively.>
- Capsada-Munsech, Q. (2017). Overeducation: Concept, theories, and empirical evidence. *Sociology Compass*, 11(10), e12518. <https://doi.org/10.1111/soc4.12518>

- Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. graduate labour market. *Economics of Education Review*, 19(2), 179–198.
[https://doi.org/10.1016/s0272-7757\(97\)00036-8](https://doi.org/10.1016/s0272-7757(97)00036-8)
- Groot, W., & Maassen van den Brink, H. (2000). Overeducation in the labor market: a meta-analysis. *Economics of Education Review*, 19(2), 149–158. [https://doi.org/10.1016/s0272-7757\(99\)00057-6](https://doi.org/10.1016/s0272-7757(99)00057-6)
- Hartog, J. (2000). Over-education and earnings: where are we, where should we go? *Economics of Education Review*, 19(2), 131–147. [https://doi.org/10.1016/s0272-7757\(99\)00050-3](https://doi.org/10.1016/s0272-7757(99)00050-3)
- Lu, Y., & Hou, F. (2020). Immigration System, Labor Market Structures, and Overeducation of High-Skilled Immigrants in the United States and Canada. *International Migration Review*, 54(4), 1072–1103. <https://doi.org/10.1177/0197918319901263>
- McGuinness, S. (2006). Overeducation in the Labour Market. *Journal of Economic Surveys*, 20(3), 387–418. <https://doi.org/10.1111/j.0950-0804.2006.00284.x>
- Meroni, E. C., & Vera-Toscano, E. (2017). The persistence of overeducation among recent graduates. *Labour Economics*, 48, 120–143. <https://doi.org/10.1016/j.labeco.2017.07.002>
- Ramos, D. (2013). Overeducation. In J. Ainsworth (Ed.), *Sociology of education: An a-to-z guide* (Vol. 1, pp. 557-558). SAGE Publications, Inc., <https://www-doi-org.ezproxy.cul.columbia.edu/10.4135/9781452276151.n296>

Sánchez-Sánchez, N., & Fernández Puente, A. C. (2020). Overeducation, persistence and unemployment in Spanish labour market. *Journal of Economic Studies*, 48(2), 449–467.
<https://doi.org/10.1108/jes-07-2019-0315>

Smith, D. (2009, February). *What is a Recession?* Federal Reserve Bank of St. Louis.
<https://files.stlouisfed.org/files/htdocs/pageone-economics/uploads/newsletter/2009/200902.pdf>

Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. *IPUMS USA: Version 11.0* [dataset]. Minneapolis, MN: IPUMS, 2021.
<https://doi.org/10.18128/D010.V11.0>

The Recession of 2007–2009: BLS Spotlight on Statistics. (2012, February). Bureau of Labor Statistics. <https://www.bls.gov/spotlight/2012/recession/>

US Census Bureau. (2012, February 1). *Educational Attainment in the United States: 2009*. The United States Census Bureau.
<https://www.census.gov/library/publications/2012/demo/p20-566.html>

US Census Bureau. (2020, March 30). *U.S. Census Bureau Releases New Educational Attainment Data*. The United States Census Bureau.
<https://www.census.gov/newsroom/press-releases/2020/educational-attainment.html>

US Census Bureau. (2021, April 20). *CPS Historical Time Series Tables* [Dataset]. Table A-2. Percent of People 25 Years and Over Who Have Completed High School or College, by

Race, Hispanic Origin and Sex: Selected years 1940 to 2020.

<https://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html>

Waldorf, B., & Yun, S. D. (2015). Labor migration and overeducation among young college graduates. *Review of Regional Research*, 36(2), 99–119. <https://doi.org/10.1007/s10037-015-0101-0>