The Probability of Being Employed in the United States

By: [redacted]

Florida International University

Department of Economics

ECO4421: Introduction to Econometrics

Dr. Alfonso Rodiguez

November 30, 2021

Abstract

The Age Discrimination in Employment Act (ADEA) of 1967, the Americans with Disabilities Act, and the Civil Rights Act of 1964 title VII are examples of the efforts taken by the United States government to provide equal opportunity of employment in the labor force. Despite these improvements, people in the United States still experience discrimination in the job market. This paper seeks to detect and analyze how social, ethnic, demographic and proficiency factors affect an individual's probability of employment in America. The data utilized to conduct this research has been extracted from the IPUMS USA microdata series, focusing on the observations pertaining to the states of Florida and Illinois, for the years 2001 and 2019. The study further examines the specific effects that the factors of sex and race have on the probability of employment in minority groups. The results from this research display that despite the implementation of anti-discriminatory policies, there are many variables that expose job seekers to discrimination in the workforce.

Keywords: probability, discrimination, employment, gender, race, demographic factors, Florida, Illinois

Table of Contents

Introduction …….. 4

Economic Theory …… 5

Description of Data ……..……. 5

Description of Methodology ……..7

Results & Analysis …… 9

Comparison Analysis ……...…… 24

Conclusion…… 24

References…… 26

Introduction

The United States is a country with a constantly growing economy. In the last ten years, the gross domestic product (GDP) annual growth rate for the United States has been an average of 3.13% (World Bank). This expansion has led to an increase in businesses and a growth in industries that have attributed to the continuous job growth in the United States. In fact, according to the Bureau of Labor Statistics, "since January 1939, job growth in the United States has averaged 2.1 percent per year" (BLS, 2016). Despite the fluctuations of job growth in individual industries due to historical shocks, new job opportunities are presented to the population. The unemployment rate, another valuable measure when analyzing the labor force, also shows this increase in opportunities. According to the World Bank, the unemployment rate has overall decreased from its peak of 6.2% in 2003 to 5.3% in 2019, before the 2020 pandemic. As the gross domestic product (GDP) expands, so does the employment rate. Steven Kapsos, an economist from the International Labour Office (ILO), found that “for every 1 percentage point of additional GDP growth, total employment has grown between 0.3 and 0.38 percentage points during the periods between 1991 and 2003”(Kapsos).

Even though there is a great amount of opportunity in the United States, not everyone in the labor force is given the same chances. Discrimination prohibits many from working to their full potential. To ameliorate this issue, during the past decades multiple policies have been implemented to prevent marginalized groups from being discriminated against in the job market. An example of this is the Age Discrimination in Employment Act (ADEA) of 1967. This act prohibits employers from making decisions to hire, fire, and promote employees based on their age (U.S. Equal Employment Opportunity Commission). In addition, the Americans with Disabilities Act (ADA) prohibits discrimination against people with disabilities and ensures they have equal opportunity in being employed (U.S. Department of Labor). Furthermore, the Civil Rights Act of 1964 title VII specifically addresses equal employment opportunity, which requires employers to hire prospective employees regardless of race, gender, and religion (U.S. Equal Employment Opportunity Commission).

Despite the advances in legislation and the progress in equity in our society, the United States is still haunted by prejudices that benefit some, but hinders others. This can lead to individuals being unemployed for long periods of time, and others being discouraged from going into the workforce altogether (Mays, Coleman & Jackson, 1996). An issue that is still prevalent today in America is the discrimination against sex. The gender wage gap is one of the main problems that affect women in America. An article published by the Pew Research Center, written by the science and social trends directors, Kim Parker and Cary Funk, states that, "One-in-ten working women say they have been passed over for the most important assignments because of their gender, compared with 5% of men" (2017, par. 5). Female workers also are at a disadvantage even before getting a job.

Another main form of discrimination is discrimination against race. Among race groups, whites compose the majority of the labor force compared to African Americans, Hispanics, Asians, and other races. In 2018, the unemployment rate was highest among American Indians

and African Americans (U.S. Bureau of Labor Statistics, 2019). Discrimination of sex and race keep minorities at a disadvantage. This leads individuals who are trying to break out of a poverty trap to struggle, when the system is mainly designed for the success of the white population, more specifically, men.

This research paper focuses on observing the probability of being employed in the United States based on individual demographics. The purpose of this study is to examine the effects of different factors, and analyze how these elements contribute to the probability of a person being employed. We examine the influence that characteristics such as race, sex, age, education, citizenship status, and many others have on employement - delving deeper into the effects of race and sex on minorities. Furthemore, our research looks at two different politically affiliated states and determines if the probability of employment is significantly different between those states. Additionally, it will provide some insight into the assumption that employers hire people based on demographic factors. If such discrimination is noted, there is a disregard to all the policies placed to prevent discrimination in the workforce.

Economic Theory & Description of Data

Prior to displaying the methods utilized in this research paper, and assessing the results encountered, it is of significant importance to define key economic concepts of employment and its related factors. This is because the study we have conducted heavily relies on employed and unemployed persons - and their distinct characteristics - in order to examine the probability of being employed in the United States. The Current Population Survey (CPS), published by the U.S. Bureau of Labor Statistics, defines the labor force as “all people age 16 and older who are classified as either employed and unemployed”. It is of interest to highlight that a person can work as little as one hour of remunerated work and still be considered a part of the employed population (BLS). When it comes to the group of unemployment, however, an individual must be actively looking for a job to be considered unemployed. Additionally, people that are not available to work, do not want to work, or have been discouraged from seeking employment for over a year, are also not a part of the labor force (BLS).

Individual characteristics that influence a person’s participation in the labor force, can determine whether they will be a part of the employed group, or the unemployed category. Demographic, social, ethnic, and proficiency factors have an effect on the probability of an unemployed person, that is seeking employment, being able to find work. For that reason, elements that have been proven to influence the average employer’s decision to hire an individual - based on the United States’ societal and economic norms - were implemented in this study. For the purpose of analyzing and predicting a person’s probability of being employed in the United States, the variables selected in this research consist of the following: race, sex, age, hispanic origin, educational level, marrital status, citizenship status, ability to speak English, presence of a disability, veteran status, and migration within the past year.

The data used to acquire the aforementioned variables came from the Integrated Public Use Microdata Series database also known as IPUMS. Such data was collected by the U.S Census Data for Social, Economic and Health Research, which is a merged dataset of sixteen

federal censuses organized by IPUMS USA (IPUMS), and by the American Community Survey (ACS), an annually ongoing survey that collects microdata on the United States' nation. IPUMS USA is a combination of the U.S Census data from 1790 to 2000, and the ACS from the year 2000 to the present year. This website consists of fundamental demographic elements at a personal and household level.

Despite the legal advances that the United States has executed to avoid discrimination in the work force, a plethora of issues pertaining to equity in the labor market remains unsolved. The probability of employment of a person can be heavily influenced by sex and race. Regulations such as the Civil Rights Act of 1964 criminalized the discrimination of a person based on sex, race, color, religion and national origin, however, there is still existing prejudices that prevent individuals belonging to a certain minority group - as women and Black individuals - from receiving equal opportunities in the work force (Blankenship, 1993, 204-205). In order for our results to be as accurate as possible, the two independent variables were included in the regression models examined. The variable *sex* was divided into male and female; and, due to the multiple categories of race, the *race* variable was separated into three independent variables - Black, pertaining to the Black/African American population; Asian, in regards to Chinese, Japanese, and other Asian or Pacific Islander persons; and Other, referring to American Indian, Alaska Native, and other races or combinations of races.

The probability of being employed is also affected by a person's age. Age is a non-linear variable that shifts from being more desirable to the workforce as a job seeker gets older from ages 16-49, to being more undesirable from ages 50 and over. This is due to the increase in productivity that is seen in the former age group, and the decrease in creativity and productivity in the latter (Skirberkk, 2003, pg.17). This pattern creates a downward facing parabola, and therefore, in our regression the age variable is represented as *age* and *age squared*. Another important independent variable added to our regression models is *disability*. This represents people with a physical disability - such as visual impairment, hearing impairment, or ambulatory difficulty - and a mental disability, which affects the self-care of an individual. A U.S. employer is less likely to hire a job seeker with a disability, as they tend to believe that such impairment can prevent the person from doing their job well. This way, disabled persons are less likely to be employed. This variable was separated into disabled and nondisabled.

Other demographic variables previously shown to impact employment status are an individual's hispanic background and their U.S. citizenship status. Hispanic men have an employment rate of 87 percent while U.S. born white men have an employment rate of 92 percent; these numbers are even lower for Hispanic women (Dunca, Hotz, & Trevejo, 2006, 232). Studies have shown that hispanic immigrants tend to have a lower educational attainment than U.S. citizens. Even if a prospective employee has the same level of education as a U.S born citizen "the foreign schooling and work experience that Hispanic immigrants bring with them transfer imperfectly to the U.S. labor market [...] U.S. employers typically place a lower value on human capital acquired abroad than on that acquired here" (Dunca, Hotz, & Trevejo, 2006, pg.232). The lower perception of knowledge creates a serious employment gap between citizens and immigrants coming to the United States, reducing the probability of employment. The *Hispanic origin* variable was divided into Hispanics and non-Hispanics, and the *citizenship* variable was categorized as citizen and non-citizen.

Marital status is also a variable that affects employment status. There are multiple different assumptions that employers have on married individuals. On one hand, single people can be seen as less committed, which employers can translate into being less committed to the job, while married employees are more likely to be open to commitment. On the other hand, other stereotypes view single people as more likely to have time for work and willing to work for longer hours, while married individuals have more obligations that take up their time. Research further shows that there is a negative relationship between marriage and female potential employees, and a positive relationship for men (Jordan, Zitek, 2012, pg. 475) . In this regression we divide *marital status* into married and not-married.

Furthermore, it is important to highlight the significance of education in regards to employment. With a greater level of education, there is a higher possibility that an individual will be employed. This is especially true today with the polarization of jobs. Due to the increase of innovation, there is a greater demand for higher level educated job seekers. Unfortunately, not everyone can afford a higher level of education, so this variable can either greatly benefit a person, or harm their chances of being employed. The *education* variable used in this research has two categories: having a Bachelor's degree or more education, or having less than a Bachelor's degree. No schooling, elementary school, middle school, secondary school, and some level of undergraduate education, all fall into the second category.

In order to analyze how different the hiring market was in the past to how it is today, we calculated regression models based on observations acquired for the years of 2019 and 2001. The year 2019 was chosen rather than the year 2020 because of the negative GDP growth rate, and the decrease of jobs created by the Coronavirus pandemic in the latter year. This way, we decided to study a more economically stable year when examining the probability of employment. Regression models were also calculated for the year 2001, to see the difference of almost two decades in the hiring markets. Besides the division of both years, we concentrated the study in two individual states, Florida and Illinois, aiming to obtain a higher accuracy rate within our models. We chose to look at the probability of employment in Florida due to our permanent residence in the state. Florida also has a large population, containing a combination of highly concentrated cities as well as less populated areas. The diversity of the state is very ample, with people from different types of backgrounds, races, and education levels. In turn, Illinois is less populous than Florida and of a different political segment. Additionally, the midwestern state has experienced an increase in urban growth, with the majority of the population residing in metro areas such as Chicago, where a lot of employment opportunities are found. Illinois has a predominantly Democratic affiliation, and was considered a "blue" state for both the presidential elections of 2000 and 2020 (270 cite); Florida followed a more conservative approach for both of the elections mentioned (270 FL). This can contribute to the states perspectives when it comes to discrimation and policies in regards to race, sex, age and more.

Description of Methodology

The methodology employed by this research paper consists of a detailed and meticulous analysis of the IPUMS USA data set extract based on different economic regression models for the years of 2001 and 2019. In order to perform such study of the data we utilized the software

Python, within the Anaconda distribution platform. Python - as defined by the Python Software Foundation (PSF) - is a "high-level programming language with dynamic semantics" (2021) and the capability of executing statiscal analysis. This program was used to examine and sort through the different variables of the aforementioned data set, as well as perform the following regression models: Linear Probability Model (LPM), Logit Model, and Probit Model. These different techniques allowed us to inspect how distinct demographic factors affect the probability of an individual being employed in the United States, more specifically, in the states of Florida and Illinois.

Before explaining our regression models, it is important to determine and characterize the variables that were assessed. Firstly, we defined our endogenous, or dependent variable as employment status, referred to as *EMPSTAT1* for the regression function. This dichotomous variable took the form of 1 or 0, where 1 means the individual is employed and 0 means the individual is unemployed. This way:

EMPSTAT1 - where 1 if employed and 0 if unemployed.

Our exogenous, or independent variables, however, were composed of both qualitative data via dichotomous variables, and quantitative variables. The binary exogenous variables were organized as the following in reference to a person:

FEMALE - where 1 if female and 0 if male;
RACEBLACK - where 1 if Black/African American and 0 if non-Black;
RACEASIAN - where 1 if Asian and 0 if non-Asian;
RACEOTHER - where 1 if American Indian, Alaska Native, or two or three major races, and 0 if not a part of the aforementioned groups;
HISPANIC - where 1 if Hispanic and 0 if non-Hispanic;
EDUCATED - where 1 if educated (Bachelor's degree or higher) and 0 if not educated (less than Bachelor's degree) ;
MARRIED - where 1 if married and 0 if not married;
NONCITIZEN - where 1 if not a U.S. citizen and 0 if a U.S. citizen;
NONSPEAKENG - where 1 if no or very low level of English and 0 if proficiency or good level of English;
MIGRATE - where 1 if moved between states in the past year and 0 if stayed in the same state;
DIFF1 - where 1 if some disability (ambulatory, self-care, vision, and/or hearing difficulties) and 0 if no disability;
VETSTAT1 - where 1 if veteran and 0 if not a veteran.

The quantitative independent variables were composed of:

AGE;
AGE2 (age squared).

Therefore, the structure of our regression model is the following:

$$EMPSTAT1 = \beta_0 + \beta_1 FEMALE + \beta_2 AGE + \beta_3 AGE2 + \beta_4 RACEBLACK + \beta_5 RACEASIAN + \beta_6 RACEOTHER + \beta_7 HISPANIC + \beta_8 EDUCATED + \beta_9 MARRIED + \beta_{10} NONCITIZEN + \beta_{11} NONSPEAKENG + \beta_{12} MIGRATE + \beta_{13} DIFF1 + \beta_{14} VETSTAT1$$

For a better comprehension of the independent and dependent variables, it is important to understand how they were classified. If a person is a Black, Hispanic, woman, not educated (has not completed a Bachelor's degree), of age 34, for example, that person would be *FEMALE=1, RACEBLACK=1, HISPANIC=1, EDUCATED=0,* and *AGE=34*. Although the above example does not involve all the variables used for this research paper, it illustrates, in a very simple manner, the utilization and format of the data. It is possible to see that dichotomous variables assume only the values of 0 or 1, and qualitative variables, such as age, do not have that limitation. Additionally, it is crucial to clarify that if a person has an indicator of 0 for *RACEBLACK,* it does not necessarily specify that such person is White; rather, it means that the individual is non-Black. Although non-Black does involve the race White, it also incorporates every other race mentioned in this study. Similarly, the same reasoning can be applied to *RACEASIAN*, and *RACEOTHER.* This way, for a person to be considered White, the base group of this research project, all three of the race variables need to be equal to 0 - *RACEBLACK=0, RACEASIAN=0,* and *RACEOTHER=0*.

We utilized the regression shown above to study the data from the years 2001 and 2019. Since the number of observations was extensively large within each of those years, we focused the calculations on two states: Florida and Illinois, as previously indicated in the section 'description of data.' Thus, the Linear Probability Model, the Logit Model, and the Probit Model, were used for the regression analyses of four sections: Florida in 2019, Florida in 2001, Illinois in 2019, and Illinois in 2001 - amounting to a total of 12 regression functions. Through the regressions computed for each of the three aforementioned models, we were able to predict the probability of employment in the United States for the different sections of our study. For that, we used the function "predict" in the Python software, which created the probability of employment for the LPM, the Logit Model, and the Probit Model. When comparing the probabilities found to the employment status used as the dependent variable for our regression (*EMPSTAT1*), we obtained the accuracy of the models mentioned above.

For each of the previously stated sections, this research paper inspects the assumption that the chosen exogenous variables are related to a person's probability of being employed in the United States. In order to understand which independent variables impact the endogenous variable we looked at the significance level. Those that had a p-value - of the t-test (LPM) and of the z-score (Probit and Logit) - lower than the significance level of 0.05 were statistically significant for our regression. This means that we were able to reject the null hypothesis that the coefficient of the variable is equal to zero at a 95% confidence level, and that the variable influences the probability of employment. If the p-value was greater than 0.05, the variable was statistically insignificant, having no effect on the former probability.

Results & Analysis

The results found in the Linear Probability, Logit, and Probit Models consisted of a combination of presumed effects, and other unexpected outcomes. Besides analyzing the

significance of the variables for the models above, we also inspected the impact of each factor in the probability of employment. This way, for the LPM, we looked at the coefficient of the exogenous variables, and for Probit and Logit, we examined the marginal effects (dy/dx) of the independent elements. The values encountered exhibited the magnitude of such factors in the probability of being employed. Referring back to the 'description of methodology' section, for our models the probability of being employed was calculated based on the employment status *EMPSTAT1*.

Florida 2019:

For the first regression, we performed the Linear Probability Model to determine the probability of employment for the year of 2019, in the state of Florida:

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	0.2562	0.0174	14.6909	0.0000	0.2220	0.2904
FEMALE	−0.1186	0.0043	−27.6704	0.0000	−0.1270	−0.1102
AGE	0.0301	0.0007	44.5221	0.0000	0.0288	0.0314
AGE2	−0.0004	0.0000	−61.0101	0.0000	−0.0004	−0.0004
RACEBLACK	0.0672	0.0071	9.4581	0.0000	0.0533	0.0811
RACEASIAN	0.0022	0.0078	0.2765	0.7822	−0.0132	0.0175
RACEOTHER	−0.0169	0.0077	−2.1900	0.0285	−0.0321	−0.0018
HISPANIC	0.0606	0.0053	11.4261	0.0000	0.0502	0.0710
EDUCATED	0.0734	0.0047	15.5685	0.0000	0.0641	0.0826
MARRIED	−0.0173	0.0046	−3.8055	0.0001	−0.0262	−0.0084
NONCITIZEN	−0.0555	0.0048	−11.4486	0.0000	−0.0649	−0.0460
NONSPEAKENG	−0.0319	0.0057	−5.5598	0.0000	−0.0432	−0.0207
MIGRATE	−0.1051	0.0158	−6.6333	0.0000	−0.1361	−0.0740
DIFF1	−0.1368	0.0074	−18.5451	0.0000	−0.1512	−0.1223
VETSTAT1	−0.0711	0.0130	−5.4706	0.0000	−0.0965	−0.0456

2019 LPM Florida Table 1.1.

As seen in Table 1.1, all the variables were statistically significant in the 95% confidence level, excluding the variable *RACEASIAN.* This element had a p-value higher than 0.05, at 0.782, displaying insignificance in impacting the probability of employment. The other factors, however, were significant in regards to analyzing employability.

The initial variable observed was the variable *FEMALE*. Our results showed that the coefficient for *FEMALE* is -0.12 or -12%. Leading to the conclusion, that, on average, being a female negatively affects the probability of being employed by 12%. Due to the prejudices against women in America, we can see that gender bias against female workers is still present in the job market.

Next, we examine the variable *AGE* and *AGE2*, which displays the relationship between a person's age and their employability. Table 1.1 indicated that the coefficient of *AGE* was 0.0301. This means that young adults have a higher chance of being employed by 3% for every additional year of age they gain. On the other hand, the variable *AGE2* had a coefficient of -0.0004, meaning that older adults face lower probability of being employed when compared to

their younger counterparts. This is in agreement with the stereotype that older people are less productive. When combining the results of these two variables we can see that as a younger individual (ages 16-50) increases in age, their probability of employment gets higher, and once they turn 50 years and older, this pattern is reversed.

Surprisingly, the variable *RACEBLACK* had a positive coefficient of 6.8%, meaning that, on average, being Black in Florida for the year of 2019, increases the probability of being employed. We did not expect to see such results for that variable, as race discrimination is still an issue in the southeastern state. Nonetheless, it is important to display the outcomes found; which could be due to a deficient data set or the accuracy level of the model. Furthermore, the variable *RACEOTHER* exhibited a coefficient of -0.0169, negatively affecting employment opportunity by 1.7%, on average, if the person is an American Indian, Alaska Native, or a mix of two or more races.

Being of Hispanic origin, represented by the variable *HISPANIC*, displayed interestings findings. Hispanic persons, in Florida for 2019, experience an increase in the probability of employment on average by 6%. As more people from Hispanic descent reside in Florida, the participation of these individuals increases in the workforce. From 2010 to 2019, Florida's Hispanic population increased by 1.4 million making it the third largest Hispanic population in the United States (Krogstad, 2020). In addition, since the number of Spanish speaking people in the state has grown, it is valuable for employers to hire people that are proficient in the language.

Similarly, the variable *EDUCATED* also had a positive coefficient. As shown in Table 1.1, an individual with a Bachelor's degree, or higher level of education, experiences an increase of 7% in the probability of being hired. As people acquire undergraduate and graduate degrees, they become more desirable to the labor force, leading to higher employment rates for those considered "Educated" in this study. While education aids in employment, marriage displays the opposite result. The factor *MARRIED* negatively influences employability by -1.7%. Therefore, it is easy to conclude that single individuals in Florida had a higher probability of having a job than the married ones.

More negative coefficients were observed for the *NONCITIZEN* and the *NONSPEAKENG* variables, assuming values of -5% and -3% respectively. This indicates that not being a citizen in the United States has a negative connotation to employability, lowering the probability of being employed by 5% when compared to U.S. citizens. Likewise, not being able to speak English reduces the possibility of getting a job by 3%. Even though Florida has experienced a growth in the Hispanic population, as previously mentioned, it can be seen from these two variables that the citizenship status and the English comprehension level of a person cannot be disregarded as they have a big influence on employment.

Additionally, the factor *MIGRATE* displayed a coefficient of -11%, demonstrating that moving between states reduces the probability percentage by 11%. Having a mental or physical disability also harms the chances of getting a job, as the coefficient of -14% for *DIFF1* demonstrated. Lastly, the variable *VETSTAT1* follows the same pattern of the former factors, diminishing the probability of being employed by 7%.

In order to evaluate the accuracy of the Linear Model we computed a confusion matrix with our true positive, true negative, false positive and false negative values. The accuracy rate is given by adding the true negative values (7802) to the true positives (21141) and dividing the sum by the total number of observations seen in Florida for the year 2019 (37323). The accuracy of this model was 77.5%, meaning that 77.5 percent of the time our model correctly predicted employed and unemployed persons.

PRED	0	1
EMPSTAT1		
0.0	7802	7015
1.0	1365	21141

2019 LPM Florida Matrix A.1

Besides the Linear Probability Model, we computed the Logit Model for the year of 2019 in the state of Florida:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	-0.1956	0.007	-29.096	0.000	-0.209	-0.182
AGE	0.0681	0.001	51.841	0.000	0.066	0.071
AGE2	-0.0008	1.36e-05	-59.444	0.000	-0.001	-0.001
RACEBLACK	0.0973	0.011	8.848	0.000	0.076	0.119
RACEASIAN	0.0064	0.012	0.543	0.587	-0.017	0.029
RACEOTHER	-0.0318	0.012	-2.753	0.006	-0.054	-0.009
HISPANIC	0.0807	0.008	9.911	0.000	0.065	0.097
EDUCATED	0.0924	0.007	12.581	0.000	0.078	0.107
MARRIED	-0.0352	0.007	-4.963	0.000	-0.049	-0.021
NONCITIZEN	-0.0866	0.007	-11.803	0.000	-0.101	-0.072
NONSPEAKENG	-0.0614	0.009	-7.005	0.000	-0.079	-0.044
MIGRATE	-0.1347	0.022	-6.060	0.000	-0.178	-0.091
DIFF1	-0.3043	0.012	-24.359	0.000	-0.329	-0.280
VETSTAT1	-0.0913	0.020	-4.466	0.000	-0.131	-0.051

2019 Logit Florida Table 1.2

In parallel to the first model, every variable was statistically significant at the level of 0.05 for the Logit analysis, with the exception of *RACEASIAN,* and *RACEOTHER.* The first race factor presented a p-value of 0.587 and the second a p-value of 0.006. Nonetheless, the remaining variables of this model were statistically significant for our regression function.

Looking at the marginal effects of *FEMALE,* presented in Table 1.2, it is possible to observe the negative coefficient of -0.20; indicating that being a woman lowers the probability of employment, on average, by 20 percent - a 8 percentage points difference from the negative 12% impact found in the previous model for that same variable. The factor *AGE* also displayed changes. In the Logit model, for every additional year of age, the probability of a person being

employed increases by 7%. *AGE2,* on the other hand, suffered only a small decrease going from -0.0004 to -0.0008.

Next*, RACEBLACK* and *HISPANIC* had an increase in influence, shifting from the coefficients of 6.8% and 6%, to the marginal effects of 9.7% and 8% respectively. This means that being Black or Hispanic in Florida, in 2019, increases a person's chances of getting a job (the reasoning concluded for such positive values can be found in the LPM description of the same variables).

In the case of *EDUCATED*, the Logit model found that having a Bachelor's degree expands the probability of employment on average by 0.092 or 9%. Being married, represented by the variable *MARRIED*, lowers the chances of employment by 3%. Not having a citizenship status (NONCITIZEN) or the capability of speaking English (NONSPEAKENG) presented marginal effects of -0.087 and -0.061, in that order. Demonstrating that each of these variables negatively affects the endogenous variable by 9% and 6% respectively - a 4 and 3 percentage point distinction from the LPM.

Ultimately*: MIGRATE* has a negative effect on the probability of employment, on average, by 13%; *DIFF1* experienced a large change from the Linear Probability Model, having a marginal effect of -30%; and veteran status - shown by *VETSTAT1* - negatively influences the possibility of employment by 9%.

The last regression for this section was the Probit Model for the state of Florida, in 2019:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	−0.1814	0.006	−29.541	0.000	−0.193	−0.169
AGE	0.0603	0.001	52.345	0.000	0.058	0.063
AGE2	−0.0007	1.17e−05	−61.761	0.000	−0.001	−0.001
RACEBLACK	0.0858	0.010	8.485	0.000	0.066	0.106
RACEASIAN	0.0031	0.011	0.282	0.778	−0.018	0.024
RACEOTHER	−0.0284	0.011	−2.650	0.008	−0.049	−0.007
HISPANIC	0.0734	0.008	9.767	0.000	0.059	0.088
EDUCATED	0.0897	0.007	13.313	0.000	0.077	0.103
MARRIED	−0.0284	0.007	−4.366	0.000	−0.041	−0.016
NONCITIZEN	−0.0786	0.007	−11.611	0.000	−0.092	−0.065
NONSPEAKENG	−0.0568	0.008	−6.974	0.000	−0.073	−0.041
MIGRATE	−0.1288	0.021	−6.088	0.000	−0.170	−0.087
DIFF1	−0.2693	0.011	−23.823	0.000	−0.291	−0.247
VETSTAT1	−0.0805	0.019	−4.279	0.000	−0.117	−0.044

2019 Probit Florida Table 1.3

For the purpose of conciseness and the simplification of the analysis, we will not expand on the results seen in the Probit Model. The outcomes of the latter were incredibly similar to the ones observed for the Logit Model, and the impacts of the variables on the probability of employment followed the same pattern. Additionally, we decided to report on the results of the Logit Model instead of the ones seen in the Probit, due to the higher accuracy found in the first.

When computing the confusion matrix for both the Logit and Probit models - like previously done in the LPM - it was possible to see that the Logit Model was the most accurate model in this section. While the accuracy rate of all three models were very similar, the Probit Model presented an accuracy level of 78.3% (calculated from Matrix A.3), and the Logit Model an accuracy level of 78.4% (calculated from Matrix A.2). This way, it is important to highlight that since the Logit Model exhibits higher accuracy, the effects of the exogenous variables on the probability of employment are more precise in that model.

	0	1
0	8920.0	5897.0
1	2179.0	20327.0

2019 Logit Florida Matrix A.2

	0	1
0	8748.0	6069.0
1	2043.0	20463.0

2019 Probit Florida Matrix A.3

Florida 2001:

From 2001 to 2019, the political, economic, and social landscape in Florida changed a considerable amount. In this section we will exhibit the results found for each of the three models in regards to the probability of being employed, for the state of Florida, in the year 2001. For such a section, we will present the tables of each model, along with the accuracy of its predictions. However, we will solely focus on analyzing the Logit Model, contrasting the results found in that model for Florida 2001, to the values seen in Logit for Florida 2019.

The first regression performed for Florida in 2001 was the Linear Probability Model:

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	0.3383	0.0360	9.3963	0.0000	0.2677	0.4088
FEMALE	−0.1699	0.0094	−18.0163	0.0000	−0.1884	−0.1514
AGE	0.0280	0.0015	19.2832	0.0000	0.0252	0.0308
AGE2	−0.0004	0.0000	−26.7604	0.0000	−0.0004	−0.0003
RACEBLACK	0.0536	0.0156	3.4384	0.0006	0.0231	0.0842
RACEASIAN	−0.0009	0.0186	−0.0470	0.9625	−0.0374	0.0357
RACEOTHER	0.0463	0.0150	3.0913	0.0020	0.0169	0.0756
HISPANIC	0.0285	0.0116	2.4623	0.0138	0.0058	0.0512
EDUCATED	0.0642	0.0114	5.6542	0.0000	0.0419	0.0865
MARRIED	−0.0254	0.0100	−2.5472	0.0109	−0.0449	−0.0058
NONCITIZEN	−0.0705	0.0102	−6.9079	0.0000	−0.0905	−0.0505
NONSPEAKENG	−0.0519	0.0122	−4.2414	0.0000	−0.0760	−0.0279
MIGRATE	−0.1204	0.0296	−4.0724	0.0000	−0.1784	−0.0625
DIFF1	−0.1383	0.0160	−8.6536	0.0000	−0.1696	−0.1070
VETSTAT1	−0.0601	0.0242	−2.4795	0.0132	−0.1076	−0.0126

2001 LPM Florida Table 2.1

From Table 2.1 it is possible to see the positive and negative effects of the significant exogenous variables in the employment probability. Like the Linear model for 2019 in Florida, the only insignificant variable found was *RACEASIAN,* with a p-value of 0.9625 - which is above the significance level 0.05. By adding the true negatives (1181) to the true positives

(4431), as shown in Matrix B.1, and dividing the number by the total observations of Florida in 2001 (8350), it is possible to see that the model had an accuracy value of 75.6%. A lower percentage in precision than when compared to the accuracy rate of the LPM of Florida in 2019.

PRED	0	1
EMPSTAT1		
0.0	1881	1685
1.0	353	4431

2001 LPM Florida Matrix B.1

After the Linear model for 2001 in Florida, we assessed the regression function for the Logit Model:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	−0.2510	0.014	−17.681	0.000	−0.279	−0.223
AGE	0.0646	0.003	23.845	0.000	0.059	0.070
AGE2	−0.0008	2.93e−05	−27.487	0.000	−0.001	−0.001
RACEBLACK	0.0618	0.023	2.705	0.007	0.017	0.107
RACEASIAN	−0.0152	0.027	−0.574	0.566	−0.067	0.037
RACEOTHER	0.0673	0.022	3.091	0.002	0.025	0.110
HISPANIC	0.0340	0.017	1.997	0.046	0.001	0.067
EDUCATED	0.0728	0.017	4.256	0.000	0.039	0.106
MARRIED	−0.0436	0.015	−2.935	0.003	−0.073	−0.014
NONCITIZEN	−0.1093	0.015	−7.267	0.000	−0.139	−0.080
NONSPEAKENG	−0.0723	0.018	−4.083	0.000	−0.107	−0.038
MIGRATE	−0.1668	0.041	−4.099	0.000	−0.246	−0.087
DIFF1	−0.2789	0.026	−10.594	0.000	−0.330	−0.227
VETSTAT1	−0.0654	0.039	−1.687	0.092	−0.141	0.011

2001 Logit Florida Table 2.2

This model had two statistically insignificant variables: *RACEASIAN*, containing a p-value of 0.566, and *VETSTAT1* with a p-value of 0.092. The results from this specific model are shown in Table 2.3, and are compared to the outcomes found for the same model in Florida for the year of 2019. Once again, we calculated the accuracy rate of the model based on the confusion Matrix B.2, and obtained a result of 75.72%.

	0	1
0	2102.0	1464.0
1	563.0	4221.0

2001 Logit Florida Matrix B.2

Finally, we performed the Probit Model for Florida in 2001:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	−0.2346	0.013	−17.893	0.000	−0.260	−0.209
AGE	0.0596	0.002	24.224	0.000	0.055	0.064
AGE2	−0.0007	2.62e−05	−28.488	0.000	−0.001	−0.001
RACEBLACK	0.0566	0.021	2.644	0.008	0.015	0.098
RACEASIAN	−0.0138	0.025	−0.555	0.579	−0.063	0.035
RACEOTHER	0.0630	0.020	3.092	0.002	0.023	0.103
HISPANIC	0.0307	0.016	1.926	0.054	−0.001	0.062
EDUCATED	0.0707	0.016	4.478	0.000	0.040	0.102
MARRIED	−0.0360	0.014	−2.602	0.009	−0.063	−0.009
NONCITIZEN	−0.1011	0.014	−7.233	0.000	−0.128	−0.074
NONSPEAKENG	−0.0666	0.017	−4.002	0.000	−0.099	−0.034
MIGRATE	−0.1575	0.039	−4.057	0.000	−0.234	−0.081
DIFF1	−0.2562	0.024	−10.467	0.000	−0.304	−0.208
VETSTAT1	−0.0517	0.035	−1.464	0.143	−0.121	0.018

2001 Probit Florida Table 2.3

In this model, the variable *RACEASIAN* is statistically insignificant, accompanied by *VETSTATS1*, and, surprisingly, *HISPANIC*. Each of these variables assume a p-value higher than the 0.05 significance level (for *RACEASIAN* the p-value is 0.579, for *VETSTATS1* it is 0.143, and for *HISPANIC* it is 0.054). The confusion Matrix B.3 analyzed for Probit, revealed an accuracy level of 75.75%. An increase in accuracy by 0.03% when compared to the Logit model for that same year, and around 0.15% when contrasted to the LPM.

	0	1
0	2088.0	1478.0
1	547.0	4237.0

2001 Probit Florida Matrix B.3

While examining our Logit Models for the years of 2019 and 2001 in Florida , we found multiple differences on the impact that distinct variables had on the probability of a person being employed. Both of the aforementioned years presented two statistically insignificant variables for the Logit Model. rThey had the element *RACEASIAN* in common, however, for 2019 the second insignificant factor was *RACEOTHER*, and for 2001, it was *VETSAT1*. It is important to highlight that, in fact, RACEASIAN was statistically insignificant for all regression models in Florida, for both the years of 2001 and 2019.

Moving on to the significant variables, it was astonishing to see how big of a negative effect being a woman had on the possibility of employment in 2001. For that year, being a woman in Florida reduced the chances of being employed by 25% - a higher impact when compared to the marginal effect of -20% seen in 2019. This could be explained by the fact that in the last decade women have been gaining a bigger participation in the labor force, as well as facing less discrimation than when compared to 2001. Additionally, *RACEBLACK* and

HISPANIC experienced an increase in percentage from 2001 to 2019, expanding from 6% to almost 10%, and 3% to 8%, respectively.

The variable *EDUCATED* also showed interesting changes. While having a Bachelor's degree in 2001 increased the probability of getting a job by 7%, in 2019, that influence grew by 2 percentage points. Which shows us that education was valued at a higher level in 2019 than in 2001. Besides educational level, the distinctions seen in marital status, citizenship status, ability to speak English, and presence of disability, exhibited important patterns. The factors *MARRIED, NONCITIZEN,* and *NONSPEAKENG,* harmed the probability of employment more in 2001 than in 2019, suffering a reduction in the magnitude of the impact. On the other hand, the variable *DIFF1* displayed opposite results, with a difference of almost 3 percentage points ranging from -27.9% in 2001, to -30.4% in 2019. Lastly, it is worth noting that the elements *AGE* and *AGE2* presented incredibly similar results for both years in the state of Florida.

Illinois 2019:

The other state that analyzed it was Illinois. Similar to Florida, we performed the previously mentioned regressions with Illinois observations for the years of 2019 and 2001.

Initially, we performed the Linear Probability Model for 2019 in the state of Illinois:

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	0.2163	0.0289	7.4786	0.0000	0.1596	0.2730
FEMALE	-0.1638	0.0070	-23.5500	0.0000	-0.1774	-0.1501
AGE	0.0342	0.0012	29.0181	0.0000	0.0319	0.0365
AGE2	-0.0004	0.0000	-37.4678	0.0000	-0.0004	-0.0004
RACEBLACK	-0.0100	0.0181	-0.5547	0.5791	-0.0455	0.0254
RACEASIAN	-0.0189	0.0090	-2.0857	0.0370	-0.0366	-0.0011
RACEOTHER	0.0362	0.0109	3.3312	0.0009	0.0149	0.0575
HISPANIC	-0.0002	0.0095	-0.0190	0.9848	-0.0189	0.0185
EDUCATED	0.0815	0.0081	10.0519	0.0000	0.0656	0.0974
MARRIED	-0.0268	0.0079	-3.3924	0.0007	-0.0423	-0.0113
NONCITIZEN	-0.0416	0.0078	-5.3485	0.0000	-0.0569	-0.0264
NONSPEAKENG	-0.0476	0.0096	-4.9389	0.0000	-0.0665	-0.0287
MIGRATE	-0.1059	0.0288	-3.6806	0.0002	-0.1623	-0.0495
DIFF1	-0.1339	0.0133	-10.0939	0.0000	-0.1599	-0.1079
VETSTAT1	-0.0745	0.0253	-2.9468	0.0032	-0.1240	-0.0249

2019 LPM Illinois Table 3.1

The results indicated that all variables, except for two, were statistically significant at the 95% confidence level. As seen in Table 3.1, *RACEBLACK* and *HISPANIC* were both statistically insignificant as their p-value is greater than 0.05, with *RACEBLACK* p-value = 0.5791 and *HISPANIC* p-value = 0.9848. Therefore the coefficients of such variables are not examined.

The first significant variable was the variable *FEMALE*. *FEMALE* has a coefficient of -0.1638 meaning that being a woman negatively affects the chances of employment by 16%. The

variable *AGE* has a coefficient of 0.0342 meaning that once young adults are able to enter the labor force the probability of being hired increases by 3.42%. On the other hand, the variable *AGE2* has a coefficient of -0.0004. From this data we can conclude that as people grow older, their chances of being employed decrease by 0.04%. Compared to our analysis of Florida, as the age of a prospective employee increases, from ages 16-49, their probabilities of employment increase, and as the person turns 50 and older, these probabilities fall. Companies are hesitant to hire older people because of the productivity stereotype, employers consider aging to be a disadvantage that makes adults less capable, less able to adapt, or less willing to go the extra mile in a work setting compared to younger adults.

In addition, the variable *RACEASIAN* also has a negative coefficient of -0.189 meaning that being Asian negatively affects the chances of employment by 1.89%. On the other hand, the *RACEOTHER* coefficient was 0.0362. This can be interpreted as individuals who identify as more than one race and/or as an American Indian experience an increased chance of being employed by 3.62%.

As expected, the variable *EDUCATED* positively increases the probability of being employed by 0.0815 or 8.15%. According to data from the U.S. Bureau of Labor Statistics, unemployment decreases as educational attainment rises. For example, "workers with a professional degree had the highest median weekly earnings and lowest unemployment rate compared to workers with a high school diploma" (Vilorio, 2016). Uniquely, the variable *MARRIED* has a negative coefficient of -0.0268. If a prospective employee is married their probability of being employed decreases by 2.68%. This can be explained by how both men and women will start planning on having a family which can lead to them being stay at home parents. Also, a married person is more likely to have more personal responsibilities and therefore less time to designate to their work, which can decrease their productivity. Due to this, businesses are more inclined to employ single people with less responsibilities.

Furthermore, both the variables *NONCITIZEN* and *NONSPEAKING* decrease the chances of being employed by similar amounts. The coefficients of both variables were -0.0416 for *NONCITIZEN* and -0.0476 for *NONSPEAKING*. It can be hard for non-citizens to obtain a formal job in the labor force because of lack of documentation such as social security number, state identification, or a passport. Non-citizens also may struggle with learning a new language which can be discouraging to employers who need workers who speak English in order to communicate with customers.

Next, the variable *MIGRATE* also decreases the probability of being employed by -0.1059, this means that just moving between states decreases the chances of employment by 11%. This can be the result of 1,766,492 (Migration Policy Institute, 2019) foreign born people entering Illinois, which we know have a harder time seeking employment in the formal sector. Equally important, the variable *DIFF1* has a negative coefficient of -0.1339. This shows that people who are disabled or suffer from mental illness have a lower probability of being employed by 13.39%, which completely disregards the Americans with Disability Act. Similarly, the variable *VETSTAT1* variable shows a negative coefficient of -0.2054 which means that a veteran's probability of being employed decreases by 20.54%. This can be explained by the fact

that veterans continue to struggle to gain employment because of cultural gaps between civilian society and their military pasts as well as the lack of effective veterans care programs in the U.S.

After looking at our results, we computed a confusion matrix to evaluate how accurate our outcomes were. This matrix is found below in Table C.1. The accuracy rate for this model was derived from adding the true negatives (2201) and the true positives (8460) and dividing that sum by the number of observations in the Illinois 2019 data. The accuracy of this model was .777, meaning that 77.7% of the time our model correctly predicted employed and unemployed persons.

PRED	0	1
EMPSTAT1		
0.0	2201	2640
1.0	419	8460

2019 LPM Illinois Matrix C.1

In addition to the Linear Probability Model, the Logit Model for Florida 2019 was also performed. Our results are as follows:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	-0.2484	0.010	-23.829	0.000	-0.269	-0.228
AGE	0.0663	0.002	30.487	0.000	0.062	0.071
AGE2	-0.0008	2.27e-05	-34.254	0.000	-0.001	-0.001
RACEBLACK	-0.0049	0.026	-0.189	0.850	-0.056	0.046
RACEASIAN	-0.0157	0.013	-1.181	0.238	-0.042	0.010
RACEOTHER	0.0432	0.016	2.752	0.006	0.012	0.074
HISPANIC	-0.0010	0.014	-0.072	0.943	-0.028	0.026
EDUCATED	0.1073	0.012	9.012	0.000	0.084	0.131
MARRIED	-0.0515	0.012	-4.371	0.000	-0.075	-0.028
NONCITIZEN	-0.0667	0.011	-5.886	0.000	-0.089	-0.044
NONSPEAKENG	-0.0677	0.014	-4.971	0.000	-0.094	-0.041
MIGRATE	-0.1470	0.038	-3.861	0.000	-0.222	-0.072
DIFF1	-0.2756	0.021	-13.316	0.000	-0.316	-0.235
VETSTAT1	-0.0912	0.038	-2.390	0.017	-0.166	-0.016

2019 Logit Illinois Table 3.2

In the Logit model, the *RACEBLACK, RACEASIAN* and *HISPANIC* variables were statistically insignificant with a z-value higher than 0.05.

Similar to the Linear Probability Model, the variable *FEMALE* has a negative coefficient of -0.2484, indicating that the probability of being employed as a woman decreases by 24.84%. The variable *AGE* has a coefficient of 0.0663 which shows that young adults are more likely to be employed. Yet, *AGE2* has a negative coefficient of -0.0008. From these results we can

conclude that as the age of an individual, of 50 years or more, increases, their probability of employment decreases by 0.08%. Next, the variable *RACEOTHER* showed that being more than one race increases the probability of being employed by 4.32%. In the same way, the variable *EDUCATED* has a coefficient of 0.1073, meaning that college graduates have a 10.73% chance of being employed. Moreover, being married decreases the probability of being employed by -0.0515 or 5.15%. Likewise, a *NONCITIZEN* and *NONSPEAKENG* both decrease the probability of being employed by 7% with *NONCITIZEN* having a coefficient of -0.0667 and *NONSPEAKENG* having a coefficient of -0.0677. In addition, the variable *MIGRATE* has a negative coefficient of -0.1470, meaning that people who move within the state have a lower chance of being employed. Similarly, *DIFF1* has a coefficient of -0.2756, meaning that the probability of being employed as a person with disabilities or mental health issues decreases by 27.65%. Lasly, the variable *VETSTAT1* has a coefficient of -0.0912, meaning that the probability of being employed as a veteran decreases by 9.12%.

We also performed a third regression for the observations of Illinois in 2019. We calculated the Probit Model, and the results are seen below:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	-0.2361	0.010	-24.384	0.000	-0.255	-0.217
AGE	0.0612	0.002	31.340	0.000	0.057	0.065
AGE2	-0.0007	2e-05	-35.945	0.000	-0.001	-0.001
RACEBLACK	-0.0062	0.024	-0.255	0.798	-0.054	0.041
RACEASIAN	-0.0143	0.012	-1.150	0.250	-0.039	0.010
RACEOTHER	0.0432	0.015	2.931	0.003	0.014	0.072
HISPANIC	-3.611e-05	0.013	-0.003	0.998	-0.026	0.026
EDUCATED	0.1047	0.011	9.397	0.000	0.083	0.127
MARRIED	-0.0406	0.011	-3.715	0.000	-0.062	-0.019
NONCITIZEN	-0.0608	0.011	-5.710	0.000	-0.082	-0.040
NONSPEAKENG	-0.0637	0.013	-4.902	0.000	-0.089	-0.038
MIGRATE	-0.1425	0.037	-3.891	0.000	-0.214	-0.071
DIFF1	-0.2507	0.019	-13.062	0.000	-0.288	-0.213
VETSTAT1	-0.0766	0.036	-2.152	0.031	-0.146	-0.007

2019 Probit Illinois Table 3.3

Due to the high rate of similarity between the 2019 Logit and Probit models, we will not expand on the Probit Model outcomes as the impacts on the probability of employment were almost exact. While comparing the models, we also saw that both Probit and Logit models have an accuracy of around 78%.

	0	1
0	2533.0	2308.0
1	735.0	8144.0

2019 Probit Illinois Matrix C.2

	0	1
0	2585.0	2256.0
1	763.0	8116.0

2019 Logit Illinois Matrix C.3

Illinois 2001:

In order to understand how much the workforce has changed throughout the years, we performed the Linear Probability Model, Probit and Logit Model on the Illinois observations corresponding to the year 2001. Similar to the Florida 2001 analysis, we will present the results for all of the models in the tables below, yet our analysis will solely focus on the Logit Model.

We initially programmed the Linear Probability Model for Illinois in the year 2001. The results of this model are presented in Table 4.1 as shown below:

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	0.3246	0.0505	6.4256	0.0000	0.2256	0.4237
FEMALE	-0.1803	0.0130	-13.9148	0.0000	-0.2057	-0.1549
AGE	0.0293	0.0022	13.4243	0.0000	0.0251	0.0336
AGE2	-0.0004	0.0000	-17.1727	0.0000	-0.0004	-0.0003
RACEBLACK	-0.0086	0.0367	-0.2351	0.8141	-0.0806	0.0633
RACEASIAN	-0.0019	0.0173	-0.1079	0.9141	-0.0358	0.0321
RACEOTHER	0.0015	0.0205	0.0751	0.9402	-0.0387	0.0418
HISPANIC	0.0473	0.0192	2.4684	0.0136	0.0097	0.0849
EDUCATED	0.0921	0.0160	5.7559	0.0000	0.0607	0.1235
MARRIED	-0.0803	0.0149	-5.3810	0.0000	-0.1095	-0.0510
NONCITIZEN	-0.0495	0.0146	-3.3921	0.0007	-0.0781	-0.0209
NONSPEAKENG	-0.0752	0.0167	-4.5006	0.0000	-0.1079	-0.0424
MIGRATE	0.0520	0.0575	0.9035	0.3663	-0.0608	0.1648
DIFF1	-0.1569	0.0251	-6.2614	0.0000	-0.2061	-0.1078
VETSTAT1	0.0196	0.0402	0.4882	0.6254	-0.0592	0.0984

2001 LPM Illinois Table 4.1

By adding the true negative and true positive and dividing it by the amount of observations of the Illinois 2001 dataset, we calculated the accuracy of the Linear Probability model. The accuracy of this model was 74%. It is also important to highlight that for the aforementioned model there was 5 insignificant variables including *RACEBLACK* with a p-value of 0.8141, *RACEASIAN* with a p-value of 0.9141, *RACEOTHER* with a p-value of 0.9402, *MIGRATE* with a p-value of 0.3663, and *VETSTAT1* with a p-value of 0.625. On the other hand, the rest of the independent variables are statistically significant with a 95% confidence interval. The specific impacts of each variable can be seen in Table 4.1 above.

PRED	0	1
EMPSTAT1		
0.0	652	998
1.0	143	2684

2001 LPM Illinois Matrix D.1

After programming the LPM, we executed the Logit Model. The results are in Table 4.2:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	-0.2384	0.017	-13.738	0.000	-0.272	-0.204
AGE	0.0585	0.004	15.181	0.000	0.051	0.066
AGE2	-0.0007	4.29e-05	-16.863	0.000	-0.001	-0.001
RACEBLACK	-0.0233	0.047	-0.494	0.622	-0.116	0.069
RACEASIAN	-0.0075	0.023	-0.324	0.746	-0.053	0.038
RACEOTHER	-0.0032	0.027	-0.119	0.905	-0.056	0.049
HISPANIC	0.0535	0.025	2.129	0.033	0.004	0.103
EDUCATED	0.1062	0.022	4.892	0.000	0.064	0.149
MARRIED	-0.1227	0.021	-5.954	0.000	-0.163	-0.082
NONCITIZEN	-0.0668	0.020	-3.408	0.001	-0.105	-0.028
NONSPEAKENG	-0.0924	0.022	-4.283	0.000	-0.135	-0.050
MIGRATE	0.0682	0.081	0.840	0.401	-0.091	0.227
DIFF1	-0.2638	0.036	-7.370	0.000	-0.334	-0.194
VETSTAT1	0.0955	0.059	1.611	0.107	-0.021	0.212

2001 Logit Illinois Table 4.2

To calculate the accuracy of the Logit Model, we input the true negative and true positive numbers into the mathematical equation that was previously explained. The accuracy for the Illinois 2001 Logit Model was 74.7%.

	0	1
0	774.0	876.0
1	255.0	2572.0

2001 Logit Illinois Matrix D.2

Similar results and accuracy rates were reported for the Probit model, as seen below:

	dy/dx	std err	z	P>|z|	[0.025	0.975]
FEMALE	-0.2250	0.016	-13.695	0.000	-0.257	-0.193
AGE	0.0554	0.004	15.528	0.000	0.048	0.062
AGE2	-0.0007	3.92e-05	-17.493	0.000	-0.001	-0.001
RACEBLACK	-0.0196	0.045	-0.431	0.666	-0.109	0.069
RACEASIAN	-0.0062	0.022	-0.280	0.779	-0.049	0.037
RACEOTHER	-0.0016	0.026	-0.063	0.950	-0.052	0.049
HISPANIC	0.0541	0.024	2.241	0.025	0.007	0.101
EDUCATED	0.1012	0.021	4.930	0.000	0.061	0.141
MARRIED	-0.1128	0.019	-5.825	0.000	-0.151	-0.075
NONCITIZEN	-0.0632	0.019	-3.396	0.001	-0.100	-0.027
NONSPEAKENG	-0.0883	0.021	-4.218	0.000	-0.129	-0.047
MIGRATE	0.0704	0.076	0.926	0.354	-0.079	0.219
DIFF1	-0.2514	0.034	-7.358	0.000	-0.318	-0.184
VETSTAT1	0.0877	0.055	1.600	0.110	-0.020	0.195

2001 Probit Illinois Table 4.3

The Probit model had close coefficients to the Logit model with some slight differences. In addition, the Probit Model has an accuracy rate of 74.6%, which is lower than the accuracy level observed for the Logit Model (74.7%).

	0	1
0	759.0	891.0
1	244.0	2583.0

2001 Probit Illinois Matrix D.3

In order to analyze the changes between the workforce in Illinois throughout the years, we reported a comparison between the Logit Models of 2001 and 2019. It is crucial that we point out the differences of the significant variables in both models. In 2019, the insignificant variables were *RACEBLACK*, *RACEASIAN*, and *HISPANIC*. In the 2001 Logit Model we observe that the insignificant variables are *RACEBLACK*, *RACEASIAN*, *RACEOTHER*, *MIGRATE* and *VETSTAT1*. To further elaborate, the variable *MIGRATE* was insignificant in the year 2001 with a coefficient of 0.0682 but for 2019 the variable was significant with a coefficient of -0.1470. Which could be explained by the increase of migrants to Illinois in the year 2019 compared to 2001. On the other hand, the *VETSTAT1* variable was insignificant for the year 2001 with a coefficient of 0.0955 but for 2019 the variable was significant with a coefficient of -0.0912. Furthermore, the variable *RACEOTHER* was also insignificant in the year 2001 with a coefficient of -0.0032 and significant for the year 2019 with a coefficient of 0.0432. Yet, the variable *HISPANIC* was significant in the year 2001 with a coefficient of 0.0535 but insignificant in 2019 with a coefficient of -0.0010. Both *RACEBLACK* and *RACEASIAN* were statistically insignificant for both years and both coefficients were showing that the probability of being employed would decrease if you were either Black or Asian.

There were also differences between the significant variables in the models. The variable *FEMALE* increases from -0.2384 in 2001 to -0.2484 in 2019. These values were both significant at the 95% confidence level.

In addition, the variable *DIFF1* had a coefficient of -0.2638 in 2001 and a coefficient of -0.2756. On the other hand, The variable *MARRIED* presented a difference in coefficients. For example, in 2001 *MARRIED* had a coefficient of -0.1227 and in 2019 the coefficient was -0.0515. This means that, compared to the past, marriage in Illinois is less important when employers hire a prospective employee. Sumarizing, not many change have occurred in Illinois from the year 2001 to 2019, which is alarming considering how much progress has been made in order to reduce discrimination in the work force.

Comparison Analysis

For a more detailed analysis on the effects of sex and race on employment, we calculated additional regressions isolating the previously mentioned variables. In order to do so, we computed different regressions for the Logit Model with the 2019 year data for the states of Florida and Illinois. We multiplied two independent variables to calculate the combined effect that these variables had on the probability of employment for certain minorities.

We first observed the impact of marriage and sex. To calculate the marginal effect of a married man on employment, we multiplied the effects of the *MARRIED* and *MALE* variables. Our results showed that, compared to a single man, in Florida, a married man is more likely to be employed by 59% (with a coefficient of 0.5944) and in Illinois this number increases to 83% (with a 0.8364 coefficient).

We executed a similar regression to analyze the probability of a married woman by multiplying the *MARRIED* and *FEMALE* dichotomous variables. As expected, our results indicated that, in both states, marriage had a negative effect on the probability of employment of a woman. Our calculations demonstrated that in Florida a married woman is 31% (-0.3175) less likely to be employed than a single woman, and in Illinois this percentage increases to 50% (-.5066). According to the former mentioned regressions, a marriage decreases a woman's possibility of being employed while this same characteristic increases the employment probability of men. Although we assumed these results, we did not expect that marriage would have such a large impact on employability. As our society has evolved throughout the years, the incoming generations tend to value marriage less. Therefore, we believed that marriage would not have as much of an impact as the results showed.

The second set of variables that were tested were Hispanic origin and sex. Due to the large percentage of Hispanic people in Florida, this was considered an important relationship to test. We found that Hispanic men (multiplying variables *HISPANIC* and *MEN*) had a positive effect on the probability of employment. Hispanic men were 62% more likely to be employed in Florida and 79% in Illinois. Hispanic women,on the other hand, were less likely to be employed. Our results show that Hispanic women are 27% (-.2795) less likely to be employed in Florida and 51% (-.5093) in Illinois. Overall, the results shown were what we expected, yet, once again, we were surprised to see the magnitude of these marginal effects on the probability of employment.

Conclusion

Inquiring from our data, we have concluded the marginal affects of demographics like sex, age, and race; focusing primilarily on Illinois and Florida for the year 2001 to 2019. Using the Linear Probability Model, Probit Model, and Logit Model we came to see to what extent every independent variable we used affected our endogenous variable. One significant factor we noticed in both states, shows that being a female in the workforce diminishes the probability of employment. Further proving the claim that there are disparities between the two genders. In addition, individuals with disabilities have also shown to have low probability of employment.

The probability of being hired for older adults has stayed negative for almost two decades. Although society has progressed greatly in terms of providing equal access to marginalized groups, the marginal effect did not decrease by a significant amount in each state when comparing the results for 2019 and 2001.

The data we have found completely disregards the laws that have been placed in order to increase equal access to minority groups. For example, the Americans with Disabilities Act, which is supposed to protect individuals with disabilities from discrimination in the work force; and the Civil Rights Act of 1964 title VII, which adresses equal employment opportunity regardless of race, gender, and religion have shown to not provide any significant help .The desired protections are evidently failing and further government actions may positively promote equality of opportunity in regards to the probabilities of acquiring employment. One possible path to mitigate these negative impacts is to shift the public outlook to disregard demographic factors, and to advocate for the compliance of these laws to be taken seriously. Although the data used was taken from a reliable database, it is worth noting that it has limitations, which can have an influence in the presence of some insignificant variables for the regressions, and in accuracy rate of under 80% for all sections of this research.

To conclude, the data has shown the significance of demographics in the probabilities of employment acquisition. Primarily, the remarkable weight that sex and race have on the decison process done by employers. The governmental actions needed are not being considered, and discussions about equality in the labor force are blatantly disregarded.

References

Blankenship, K. M. (1993). Bringing gender and race in: US employment discrimination policy. Gender & Society, 7(2), 204-226. https://journals.sagepub.com/doi/pdf/10.1177/089124393007002004?casa_token=C4iLXrX9lMYAAAAA:Hyr2hf9t2Zg3wNUBASSJGXG0dPlQ8CDPFuMCXW8SVxZKKEOz7aWrWAQc4eJ2aTwkdGVDzzEn89uS

Bureau, U. S. C. (2021, November 17). American Community Survey (ACS). Census.gov. Retrieved November 29, 2021, from https://www.census.gov/programs-surveys/acs.

Duncan B, Hotz VJ, Trejo SJ. (2006) Hispanics in the U.S. Labor Market. In: National Research Council (US) Panel on Hispanics in the United States; Tienda M, Mitchell F, editors. Hispanics and the Future of America. Washington (DC): National Academies Press (US); . 7. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19908/

Illinois Department of Central Management Services. (2019, April 26). *Asian American Employment Plan*.

Impacts of Latino Population Growth on Comprehensive Planning in the Chicago Metropolitan Region. (2008, December).

Jordan, A. H., & Zitek, E. M. (2012). Marital status bias in perceptions of employees. Basic and Applied Social Psychology, 34(5), 474–481. https://doi.org/10.1080/01973533.2012.711687

Krogstad, J. M. (2020, July 10). *Hispanics have accounted for more than half of total U.S. population growth since 2010*. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/07/10/hispanics-have-accounted-for-more-than-half-of-total-u-s-population-growth-since-2010/

Loughran, D. S. (n.d.). *Why is Veteran Unemployment so High?* https://www.rand.org/pubs/research_reports/RR284.html

Mays, V. M., Coleman, L. M., & Jackson, J. S. (1996). Perceived race-based discrimination, employment status, and job stress in a national sample of black women: implications for health outcomes. Journal of occupational health psychology, 1(3), 319. https://psycnet.apa.org/fulltext/1996-04828-006.pdf

Migration Policy Institute. (2019). *State Immigration Data Profiles* [Demographics & Social for Illinois]. https://www.migrationpolicy.org/data/state-profiles/state/demographics/IL

Python. (2021, November 19). *What is Python? Executive Summary*. https://www.python.org/doc/essays/blurb/

Skirbekk, V. (2004). Age and individual productivity: A literature survey. Vienna yearbook of population research, 133-153. https://www.demogr.mpg.de/papers/working/wp-2003-028.pdf

Team, M. P. C. U. X. U. I. (n.d.). U.S. Census data for social, economic, and Health Research. IPUMS USA. Retrieved November 29, 2021, from https://usa.ipums.org/usa/.

270 to Win. (n.d.). *Florida*. https://www.270towin.com/states/Florida.

270 to Win. (n.d.). *Illinois*. https://www.270towin.com/states/Illinois

GDP growth (annual %) - United states. World Bank. (2020). Retrieved November 29, 2021, from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=US.

U.S. Bureau of Labor Statistics. (2016, August 1). Current Employment Statistics Survey: 100 Years of employment, hours, and earnings : Monthly Labor Review. U.S. Bureau of Labor Statistics. Retrieved November 29, 2021, from https://www.bls.gov/opub/mlr/2016/article/current-employment-statistics-survey-100-years-of-employment-hours-and-earnings.htm.

U.S. Bureau of Labor Statistics. (2019, October). *Labor Force characteristics by race and ethnicity, 2018*. https://www.bls.gov/opub/reports/race-and-ethnicity/2018/home.htm

Vilorio, D. (2016, March). *Education Matters*. U.S. Bureau of Labor Statistics. https://www.bls.gov/careeroutlook/2016/data-on-display/education-matters.htm