**“How Does The Labor Force Participation Changes In The United States”**

**Professor Name :** Kevin Foster.

**Course Name :** Statistics and Introduction to Econometrics.

**Institution :** CUNY , The City College Of New York .

[Colin Powell School for Civic and Global Leadership](https://www.ccny.cuny.edu/colinpowellschool).

**Student Name :** Amira Elmakawy.

Md Muhibul Islam.

**Due Date :** Dec/18/2022

**Introduction :**

This paper discusses the different factors and forces behind the change and cyclicality in the labor force participation in the U.S , which is mainly associated with the economic activities.

The paper studies a combination of two parts , which are the main conclusions from the previous related research papers and the main results of our data analysis and hypothesis testing on the two datasets that we used here which are the “American Census Survey “ ( *ACS 2017 ny\_data*) that was used several times during our semester and which is the first data used for this paper which is mainly a demographic large questionnaire through different ways such as telephone interviews , mails and visits to about 3.5 million households annually .

Besides the “ National Health Survey” (Household\_Pulse\_data) (*NHIS\_2021*) data which does summary household health statistics using their questionnaires and interviews and uses these questionnaires as a source information on the health of the population with Health data statistic in the United States which collects information related to the U.S households around the 50 states and classifying them using their demographics and some other factors like their vaccination status , their family sizes, Their psychological status , employment status and more.

Participation rate in the labor force basically includes both percentages of the population aged 16 and older who are already enrolled in the labor market and the others who are unemployed but seeking to find and get a job. While here , There’ll be two subsets of people aged (55-79) to those aged (25-55) , which shows that the difference in their means of (*1.45*) for the former , to (1.81) for the latter as a result of the slow growth in population and the rise in the retirement rate for baby boomers especially for the next years.

In general , **The main result that is highlighted in both our data analysis and the other research papers that we used is the observable gender difference distribution between males and females in the participation of the labor force** , and mostly between people aged (25-54) which are called “*prime-age workers*” .

This paper also discusses how the Covid-19 pandemic strongly affected employment level by causing huge changes all over the economy .

The interesting issue about this topic is that there’s a decline in the prime-age workers too which in the short-run and long-run leads to negative effects in the Economic growth , Income , Standards of living for individuals and their well-being.

The more interesting result also is that despite the large drop in participation in the labor market and the tending to stay home and feel the risk in the working place as a result of high contacting , there is a gradual relief started to happen nowadays and the trend to work in the regular working environment started to emerge again instead of preferring not to work at all or to work from home which is called the Recovery stage.

**Literature Review:**

We’ve reviewed some papers that are used as a reference and related to our topic , We used almost four or five papers in our research , The first three papers are related to the behavior of labor force participation in general , and the other two are related to the COVID-19 impact on this labor force participation behavior.

The first paper is **“ The Changing Cyclicality Of Labor Force Participation “** by (Willem Van Zandweghe” , 2017 ) , That reviews how the employment level participation behavior rate can be **countercyclical** or **procyclical** , the data used in this paper was acquired from National Bureau of Economic Research , As it also studies that how more “cyclical” the rate of participation between the *prime-age workers* versus with the older workers whose participation rate are becoming “counter-cyclical” over time .

This paper also figured out that over time , The labor force participation rate of prime age workers has become more “procyclical” than the rate for older workers which had been switched to “counter-cyclical” , and this was something similar to what we did in our statistics thatwe found using the ACS data, as the participation mean for older workers is getting lower than participation rate mean for prime age workers , It also shows and discusses the volatility and how productivity of labor growth differs and changes according to any negative shocks , which may be such as the impact of Covid-19 which is considered as a negative shock on this side without even mentioning that directly in the paper as this paper was before the pandemic in 2017.

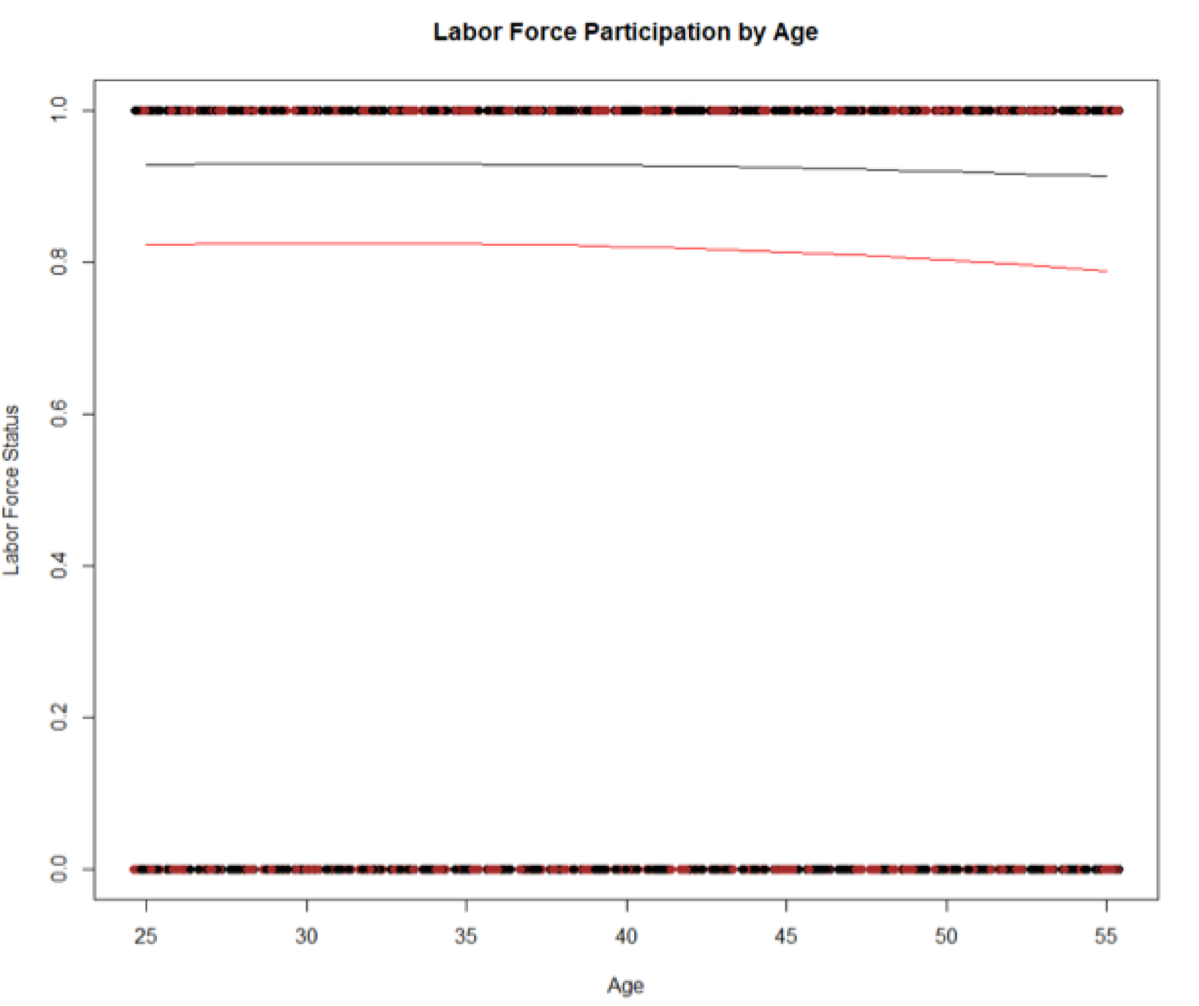
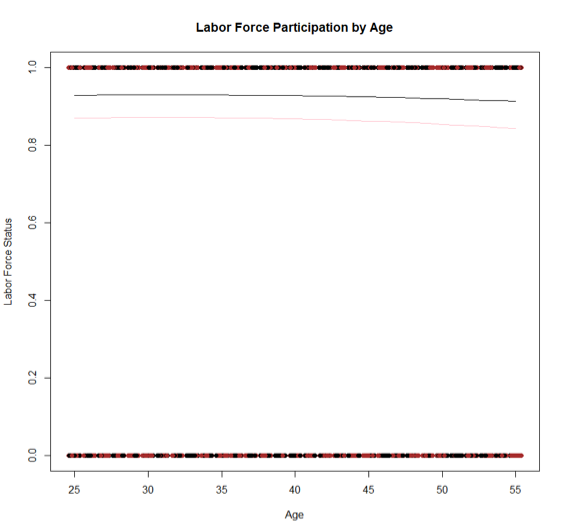
The second paper is **“The Effect Of Women’s Labor Force Participation On The Distribution Of Income In The United States”** by [Judith Treas].

Using public use samples from decennial censuses and some tables from the **CPS** which studies and discusses that there could be a decline in family income inequality and this might lead to an equal distribution between families for husband-wife families than otherwise to be . or at least that helps create inequality to be to a fair-extent stable.

This paper also reflects the share of husbands and wives in the labor market where the working wives reduced the burden of the unemployment problem in the 1980s , and it also might raise demand in houses in the housing market as she helps the husband in home-buying by using the logistic model of home purchases which lead to an improvement in the economy.

For our analysis ,We found out that the “married woman” is more likely to be in the labor force market than the “unmarried woman” although both are there but the married woman has a bigger impact on the employment level .

The figure below from our analysis using the ACS data indicates the participation of both married males and females where the gray line is for married males and the pink line is for married females (Left Figure) , as we can see that both rates are very close to each other at a high labor force status with a higher mens’ labor force participation.



(labor force for married women) (labor force for unmarried women)

From our analysis , the mean percentage for the “never married woman” is (81%)

While the mean percentage for “married women” is (86%) . So it’s a little higher than the “never married woman” (Right Figure) , So our analysis in this part has been compatible with this paper’s results .

The third paper is :  **“WHAT WE KNOW AND DON’T KNOW ABOUT DECLINING LABOR FORCE PARTICIPATION”** by Eleanor Krause and Isabel Sawhill . The Brookings Institution , May 2017.

This paper studies the decline in the labor force participation for the prime-age workers , and also discussesthe reason for this decline and how the supply of and demand for labor declines thanks to the automation and technological changes that created some gaps in workers wages and skills lack , plus the other factor of illness and poor health and high mortality levels.

These are all reasons for the employment level dropping PLUS the preference of some persons to receive unemployment benefits rather than working and this happens especially between men.

According to these consequences of the labor force participation drop , There'll be a negative effect on the person's mental health , , marriage decline and may cause the earlier rates of death , all of this leads to the higher poverty rates.

This paper also concludes that although the improvements and incentives regarding the womens’ workforce laws that can encourage women towards the workforce , this paper shows in spite of all of that **,** there’s still a decline in women participation in the labor force.

Regarding the part of education , It concludes that the highest participation rate is coming from workers of 4 years of college or greater which is something that our paper concluded too using the (ACS\_2017 data) .where in our data (16452) out of (30802) persons with 4 years education participated in the labor force in our sample, where in this *Eleanor Krause and Isabel Sawhill* paper , It ensured that using the “CPS data” where the less educated groups don’t receive the essential trainings to keep up with today’s available current jobs as they are less-skilled and now everything depends on the automation and is used everywhere , plus the disability or illness , in addition to the Covid-19 pandemic nowadays, as health status is getting worse .

So *Eleanor Krause and Isabel Sawhill*paper contains basic and main statistics such as finding the means for men with different levels of education to check their participation in the labor force , and this is actually something that also we did.

There are more research papers used in order to check this labor force participation behavior subject to the COVID-19 pandemic such as :

**“Effects of the COVID-19 Recession on the US Labor Market: Occupation, Family, and Gender”** by Stefania Albanesi and Jiyeon Kim , published in 2019.

This paper might be the longest paper that we’ve recently studied in order to see how the workers behavior changed after the pandemic of COVID-19 with respect to their occupations , genders ,and family status . So it might take a larger proportion in this Literature Review.

This paper said that COVID-19 recession led to employment losses and they were larger for women than men (which we also found already using our NHIS\_2021 data) and the reason for this was according to both sides , Demand Side and Supply Side .

Demand Side indicated differences by gender in the distributions of occupations , in other words they have differences in their occupations' nature in which men are concentrated more in the goods production industries and occupations while women are concentrated more in services and professional occupations . where the nature of mens’ occupations seemed to be more cyclical and have more risks than womens’. But in the COVID-19 times the employment cyclicality actually affected the women more as people were afraid of using social services and any other kinds of services because of the COVID infection which led to the decrease in females in employment , and that was something that we found using the NHIS data 2021 too.

This paper also concluded that “Married women” were more likely to lose their jobs during COVID-19 pandemic as they had to stay home more and take care of their kids as also the in-person child care declined in addition to that the schools closed at this time of pandemic , So this made too many people to quit the labor force and this part of the problem argues the Supply Side of labor force.

This paper also argues an important thing about the COVID-19 pandemic where it compares between the Great Recession and the Covid-19 Pandemic Recession and how they affected employment .

By using Regression analysis , They found that there are two huge flows in economy which are “ Employment to non-participation “ which presents the Supply Side , and “Employment to unemployment “ which presents the Demand Side , and by checking that and concluding the results , they used some factors such as family status , occupation , and gender in order to tell about the jobless phenomena and its recovery and how can remote work affect on this employment status and gender gap.

They mentioned that in general woman doesn’t leave the employment very easily in regular recessions , as she tries to help and support her family facing the threat of her husband job loss , while this paper studied how women left the labor force in the COVID-19 recession by a larger percentage than men , and it also studied the effect of this COVID-19 in its most vital and severe period where it was a tough time “at the beginning of it” until November 2020 using the Current Population Survey data , While our paper used the 2021 National Survey Interview Survey Data which was most recent .

This paper showed that the decline in men employment was more than the decline in womens’ employment in the Great Recession , while the opposite happened in the COVID pandemic recession where women’s employment decline was more than mens’ and that was for each family group .

So the high fraction of women participation in the services occupations reflected a high fraction of losing these occupations in the COVID-19 pandemic times due to the infection risk , in addition to the decline in schools activities and the in-person child care services , so also that’s why women employment declined more than mens’ as mentioned before.

Another important thing that this paper highlighted is the disproportions of job loss between lower wage workers , and this was another factor that might led to the womens’ decline in labor force participation as there’s “Child Penalty” that gives the women with no children less wages and that can be a cause for women to lose their jobs specially in the pandemic period as she sees that it’s unfair to her as a woman with no children to get paid less than the woman with children specially in a hard times like the pandemic period.

This paper classified kinds of occupations with respect to flexibility and degree of contact where the flexible ones can be carried out and can be low-contact remotely , while the non-flexible ones are the outdoor activities which describe working on sites with a high contact and cannot be carried out remotely. Whereas the inflexible/high contact occupations are the most affected group which is represented by the majority of females participation in the services occupation , In contrast , men who usually work in a low-contact flexible and non flexible environments compared to women.

The second worst group of occupations was the inflexible low-contact group that witnessed a large decline in its occupations .

Stefania Albanesi and Jiyeon Kim paper described how the gender plays an important role in the jobs and occupations distributions , where females share the most in the occupations which are flexible and have high contact , in other words these jobs can be easily done remotely without contact , while the opposite of these jobs from the inflexible high contact jobs that witnessed the highest decline in the employment followed by the inflexible low contact jobs in April 2020 .

While the flexible low contact jobs witnessed the lowest drop in employment as it was the most preferable kind of work at these times where the drop for women and men was almost the same but still the women drop was slightly more compared to men.

One pattern that appeared in this period of COVID recession was that the female employment recovery was less than mens’ in flexible low contact jobs where these kinds of work used to have the largest percentage of women which had been largely affected by the COVID hit .

On the other side of the other kinds of occupations which are the inflexible ones , The drop of mens’ employment was higher than the womens’ that was because men used to have the highest percentage of participation in labor force in these kinds of jobs including the essential frontline workers which came up with a bigger decline in employment due to the COVID pandemic.

In this case we’ll consider the jobs which are inflexible with high contact were the most cyclical jobs in the Covid period , and in contrast the flexible with low contact jobs were the least cyclical jobs in the COVID period . In addition to that some workers who used to work in the inflexible high contact jobs did a shift and switch in their careers into the low contact flexible jobs such as workers who used to work in goods production industries tended to shift into the service provider jobs.

Employment behavior over the period of the pandemic was determined by the demand and supply elements and components , such as the demographic factors like age , gender , family status , in addition to the acquirement of education and children , In other words , since the Supply and Demand factors affected on the employment participation in the labor force , It also affected on its decline for both genders , especially the women , from these factors that affected women decline in the labor force is being married and having children which was considered as a Supply factor , and it almost contributed by two thirds of the employment decline , besides the changes in occupations which contributed by the rest of this percentage share.

In general , there’s a large gap in drop between women and men were always the decline in women labor force participation is higher than men specialty if women are married and have children , That reduced her employment share by larger percentage than those married women with no children , but still the decline in women's participation in labor force is higher than the mens’ and that was because of the difference in distribution of different kinds of jobs specially for parents than the single ones.

One of the Demand Side elements that changed the employment status to unemployment was the destruction of jobs which badly appeared in the beginning of the COVID-19 recession .

While one of the Supply Side elements that changed the employment status to unemployment status is that the employed workers voluntarily tend to leave their jobs causing this nonparticipation of workers in the labor market which created the phenomena of the employment to nonparticipation which mainly caused a shift in the Supply Side down with gender differences and this was largely contributed by women especially these single women with children followed by married women without children compared to men who usually did a shift from job to another job , So in general women affected largely on the unemployment and its duration in the workforce in the U.S and that was since the early 1990s (Abraham and Shimer 2001).

**# Recovery :**

Although women were the one who largely contributed to the unemployment phenomena , they became also the ones who largely affected the employment recovery , as they started to join the labor force again. Another reason for the employment recovery was that because as mentioned before too many workers left the goods-production jobs which were inflexible with high contact , afterwards , these kinds of jobs started to be filled up again by different workers.

Another factor of recovery was that because of the Automation and technological progress phenomena that showed up after the COVID pandemic appearance which encouraged some workers to go back to work after they were afraid of the COVID infection in the working place although that the automation on the other hand created a skills and wages gap between workers.

The second paper that’s related to the COVID-19 impact on the labor force participation and which typically ensured the results of the first paper is :

# **“U.S. labor market inches back from the COVID-19 shock, but recovery is far from complete”** BY [RAKESH KOCHHAR](https://www.pewresearch.org/staff/rakesh-kochhar) AND [JESSE BENNETT](https://www.pewresearch.org/staff/jesse-bennett).

This paper says that The US labor market is trying to recover from COVID-19 but still far away.

As the worldwide coronavirus outbreak started in Feb 2020 which had a dramatic impact on the US labor market which led to a record unemployment rate. Lot of employees left their jobs at that time. On Feb 21 employment was a lot less than in Feb 2020. So it will take more than three years to recover from the positive scenario which is likely to occur from 2018 to 2019. There are some of the noticeable COVID-19 effects on some Americans more than others . Women have been affected more than men. Uneducated, less educated people, young adults, immigrants as well as Hispanic and non hispanic people faced a lot of job losses .

Some key points about how worldwide recession is affecting labor force participation and unemployment among workers a year in the USA are : Women are more likely than men to quit employment in the first year of the COVID-19 recession , From 2020 to 2021, Women accounted for a majority of the decrease in the labor force in the first year of the downturn although they make up less than half the U.S. workforce. Apart from that, For women, the labor force participation rate in February 2021 was less than a year earlier. The decrease in the labor force participation rate for workers overall – exceeded that seen in the Great Recession , according to Bureau of Labor Statistics data. Although lower than a year ago, the labor force participation rate has risen in recent months. The rate for women had fallen in April 2020, and the rate for men had dropped in the same month. Since then, the recovery appears for women sharper than for men which is something ensured by the first paper too.

Black and Hispanic women accounted for much of the decrease in labor force participation among women where both Hispanic and Black women accounted for 46% of the total decrease in labor force participation among women.While in our NHIS data , The analysis told us that the high work loss was mostly for the non hispanic people than the hispanic ones ,

Low-wage workers' jobs experienced the greatest drop in employment followed by middle-wage workers, while in contrast high-wage workers employment was roughly unchanged , and we explained this by doing some summary statistics in the NHIS data for the Income variable and its relation to the work loss variable.

**Means(simple graphs, correlations, differences of means) :**

**By starting with the ACS\_data 2017 :**

We’re using the LABFORCE (labor force ) as the dependent variable here.

* We Used the subset of prime-age workers (25-55) years old. By finding the mean and median of this subset participated in the labor force , we found that the mean is 1.816 where the median is 2.000 .

| **LABOR FORCE** |
| --- |
| Min. :1.000 |
| 1st Qu.:2.000 |
| **Median :2.000** |
| **Mean :1.816** |
| 3rd Qu.:2.000 |
| Max. :2.000 |

And by doing Summary statistics for the Labor Force and Education levels , where the value of 1== Not the labor force . 2== Is in the labor force. We got these results

**N/A or no schooling**  **Nursery school to grade 4** **Grade 5, 6, 7, or 8**

1408 283 1564

**Grade 9 Grade 10 Grade 11**

857 1073 1280

**Grade 12 1 year of college 2 years of college**

23018 8251 7673

**3 years of college 4 years of college 5+ years of college**

0 18441 13534

These results indicate that the majority of workers are subject to the educational levels of those who are in Grade 12 and who acquire education of 4 years of college. And followed by the 5+ years of college.

By specifying this in a table , that ensures the same result that we mentioned before of the segment share of each educational level in the labor force participation.

**1 2**

**N/A or no schooling** 725 683

**Nursery school to grade 4** 105 178

**Grade 5, 6, 7, or 8** 589 975

**Grade 9** 358 499

**Grade 10** 535 538

**Grade 11** 581 699

**Grade 12** 5717 17301

**1 year of college** 1519 6732

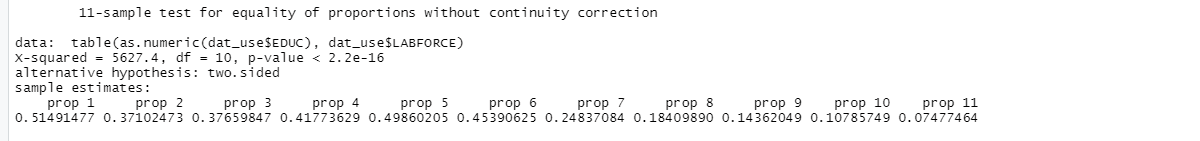
**2 years of college** 1102 6571

**3 years of college** 0 0

**4 years of college** 1989 **16452**

**5+ years of college** 1012 12522

And by doing proportions of these statistics we’ve got these proportions :



So again these results show that the majority of labor force participation is from advanced education and Grade12 and bachelors degree and there’s a strong relation between education and the employment levels as p-value is less than significance level so we can reject the null hypothesis of no relation between both factors.

* While doing summary statistics for people who are in the workforce , the majority is enrolled in the Labor force, and there are 63,150 persons who joined the labor force while only 14,232 are not in the labor force in this subgroup.

So the majority of this subset is still in the labor force.

Mode **LABFORCE=1** **LABFORCE=2**

logical 14232 63150 out of total 77,382

**By using the NHIS\_2021 data :**

We’ve used the WRKLOSS (Work Loss) variable as a dependent variable .

We used the subset of people or workers whose ages range between (25-55) years who presented the *Prime-Age Workers* in the labor market.

By doing Summary Statistics for variables that were used in the next stage in analysis : **Summary Statistics for variables :**

The HAD COVID variable : We found that in our sample the majority of it are people who did not have the Covid , while the minority were those who the doctor told them they had Covid.

The INCOME variable : (Even if we couldn’t use in the regression after ) but we did some summary for it and we found out that majority of our sample with a high income range between ( $ 100 K - $ 149 K) followed by those group with income range between ( $ 50 K - $ 49 K ) .

The Workloss variable : the majority of people who are those with no job loss recently (59095) while the rest which was a small fraction of the sample (8058) recently lost their jobs.

The vaccination receiving variable (RECVACC) : Most people got vaccinated while the rest did not .

yes got vaxx no did not get vaxx

60326 7937

We’ve added a variable such as (WORRY) as the COVID-19 pandemic affected also on the mental health of some people so we thought that the worry will have an effect on the labor work loss , and we found from this 2021 data sample that almost (17936) people in this sample have worry . while doing a table between the WORRY variable and the WORKLOSS , we found that (2493) people who lost their jobs had several days of worry over past 2 weeks , however (1740) people of nearly everyday worry recently lost their jobs and (1891) people who never worried over the past 2 weeks recently lost their jobs , while on the other side , a high fraction of people (15409) of “Several days worried over the past 2 weeks) recently have not lost their jobs.

The ANXIOUS variable which is similar to the WORRY variable , we’ve got (19794) people have Several days anxiety over the past 2 weeks compared to (26611) who were with no anxiety.

While doing a table between Workloss and ANXIOUS , we found that the majority of people who recently lost their jobs had several days of anxiety over the past 2 weeks , followed by nearly everyday anxiety .

**no anxiety over past 2 wks**

26611

**several days anxiety over past 2 wks**

19794

**more than half the days anxiety over past 2 wks**

6140

**nearly every day anxiety**

8623

We were trying to add the Working on site in the model , but we’ve got (0) results for working on site compared to (27846) which appeared to be not beneficial results so we excluded this variable from the model .

So we added the WORKING REMOTE Variable as its strongly related to our topic of the COVID-19 pandemic , and by doing a summary , we’ve got that (22863) persons work remotely compared to (38229) persons who don’t , So there is no big difference between both .

While doing a table between the Workloss Variable and the Work remotely variable we’ve got that the majority of people who recently lost their jobs don’t work remotely (5157) compared to (2055) recently lost their jobs while working remotely .

worked remotely no

no recent HH job loss 20787 32995

yes recent HH job loss 2055 5157

For Income Variable , while doing table between Income and Workloss , The majority of people who recently lost their jobs (1438) people have (HH Income less than $ 25 K) followed by (1053) people have (HH Income $ 50 K - $ 74.9 K) , And then (930) people have (HH Income $ 25 K - $34.9 K ).

For the KINDWORK (Kind of work variable ) , Majority of people work in a private company (21370) , followed by those who work in the Government ( 6378) and then the Self Employed people ( 4966) .

While doing a table between the Workloss and the Kind of work variable , We found that the majority of people who recently lost their jobs were those people who used to work in a private company , followed by the Self-Employed ones .

For the Gender Variable , We found that the majority of people in our sample are females and also who recently lost their jobs are mainly females than males .

female male

no recent HH job loss 34678 23230

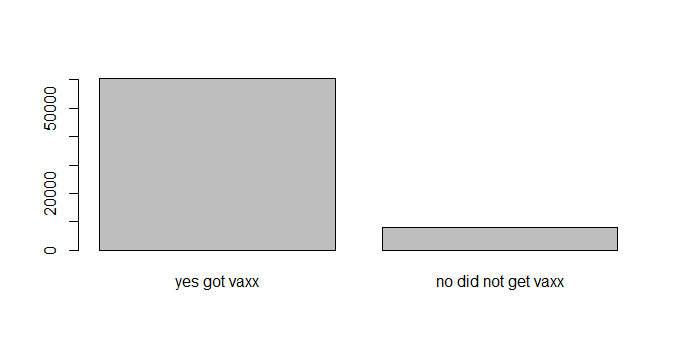
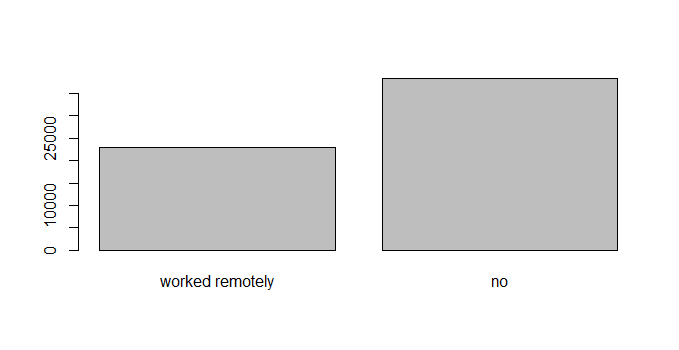
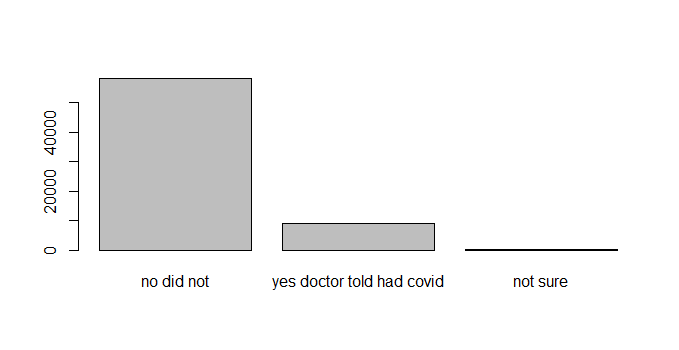
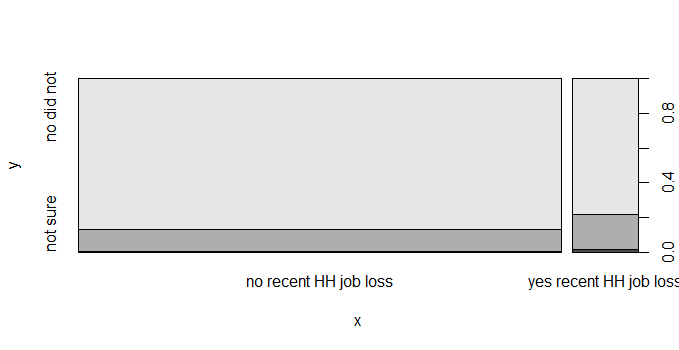
yes recent HH job loss 4751 3057

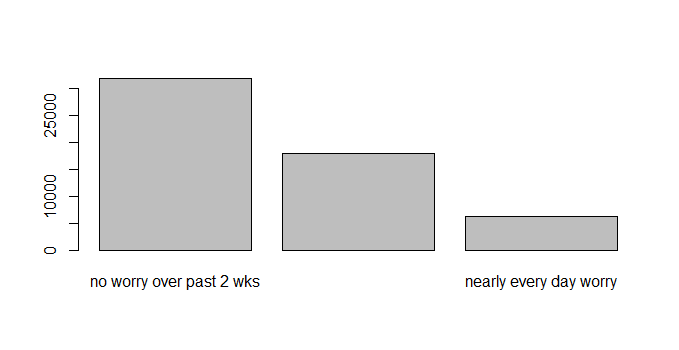
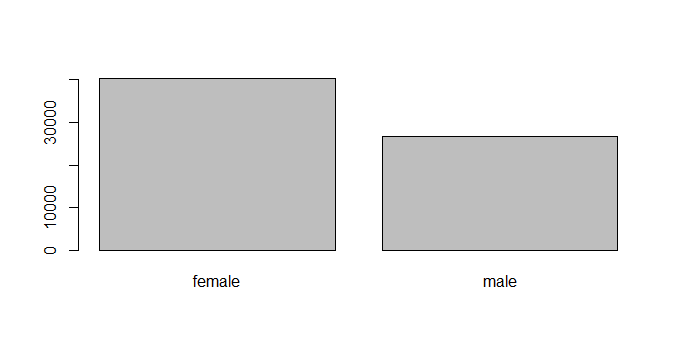
By summarizing the Education Variable , we found out that the majority of sample has a Bachelor's Degree (20075) and this is something that we indicated too in the the ACS\_2017 data followed by the “Advanced Degree” major , and while doing a table between the workloss and Education , We found that majority of people who recently lost their jobs in 2021 are from the “some college” major (2129) , followed by “Bach Degree” (1939), and then the “Adv Degree” (1405) .

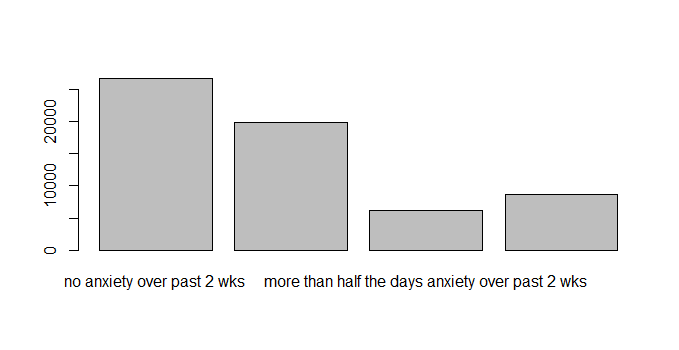
And for the RRACE (Race) Variable , we found that most of the people in the sample are White , followed by Asian , and then Black . And adding a table of this variable with the workloss , We’ve got that the majority of those who recently lost their jobs are from the White people , followed by Black , and then the Asian .

While adding the Hispanic Variable , We’ve got that the majority of those who recently lost their jobs are not hispanic while we expected the opposite . But this can be normal because the not hispanic people in the sample are more than the hispanic ones.

**Basic Summary Statistics graphs for NHIS\_2021 data for different variables :**

****

****

****

**Simple Regression :**

**For the ACS\_2017 data :**

When using the proportion tests , The model is***statistically significant*** where education is considered as an important factor for the employment level as p-value is "2.2e-16" which is less than the significance level of 0.1 .

While doing linear regression : using the **LABOR FORCE** variable to be the dependent variable where **age** , **marital status** , **female** , **race** to be independent variables in the model , while the null hypothesis is that there is no relation between labor force and those independent variables , and instead , The alternative hypothesis is that there is a relation between labor force and those independent variables .

So after doing the **Simple Linear Regression** , we found that the model is Statistically significant as the p-value is so small and less than significance level , R squared is small too, and F-Statistic is high which means that there’s a high dispersion of the data scattered from the mean .

lm(formula = LABFORCE ~ EDUC + female + AGE + MARST + RACE, data = dat\_use)

“Simple Linear Regression”

Residual standard error: 0.3691 on 77356 degrees of freedom

Multiple R-squared: 0.09249, Adjusted R-squared: 0.09219

F-statistic: **315.3** on 25 and 77356 DF, **p-value: < 2.2e-16**

**Residuals:**

**Min 1Q Median 3Q Max**

-1.03612 0.03343 0.12564 0.20227 0.65391

And the **Simple Linear** **Regression Equation** will be :

LABFORCE = 1.6080269 + 2.193 EDUC + -0.0908451 female + -0.0011445 AGE + -0.1824 MARST + -0.2449 RACE

So according to these results , *we can reject the null hypothesis of no relation* and conclude that there is a relation between these main demographic variables and the labor force.

By doing **another model** while adding the **marital status** variable plus **marital status squared** , the model is still statistically significant at 95% confidence level .

model\_2 <- lm(LABFORCE ~ AGE + EDUC+ female + + I(MARST)^2 + MARST +RACE,data = dat\_use)

Except for some **Categories in the marital status** factor such as "MARSTdivorced , MARSTmarried spouse absent , MARSTdivorced ,MARSTwidowed , in addition to EduGrade10 with some types of Race.

By adding the AGE squared and cubed to the model , The model is still unchanged and it will be **significant**.

model\_3 <- lm(LABFORCE ~ AGE + **I(AGE)^2** + **I(AGE)^3** + EDUC+ female + **I(MARST)^2** + MARST +RACE,data = dat\_use)

**For the NHIS\_2021 data :**

We used the (WRKLOSSRV) Workloss variable as a dependent variable while the independent variables are (WORRY , HAD COVID , ANXIOUS , RECVACC , KINDWORK , GENID\_DESCRIBE)

Unfortunately , We could not do the linear regression as we tried so much , So we had to use the non-linear regression model using the logit and probit methods.

**Non-linear Regression model :**

**For the ACS\_2017 data :**

By doing some Interactions :

The model will look like these two formulas :

model\_4 <- lm(LABFORCE ~ AGE + I(EDUC)^2 + EDUC\*MARST + female + MARST + RACE,data = dat\_use)

model\_5 <- lm(LABFORCE ~ AGE + I(AGE^2) + EDUC+ female + + I(MARST)^2 + MARST +RACE,data = dat\_use)

With similar results while the model is still significant , So in this case again we can reject the null hypothesis of no relation between Labor force participation and ( AGE , EDUCATION , RACE , MARITAL STATUS , AND BEING A FEMALE) and Labor Force Participation dependent positively on the EDUCATION. So education has to be a main factor to be considered in this relation.

By doing the ***logit model*** That will be the equation :

model\_logit1 <- glm(IN\_LABFORCE ~ AGE + I(AGE^2) + EDUC+ female + MARST +RACE , family = binomial, data = dat\_use)

And by comparing the labor force between males and females as mentioned above , We’ll figure out **that participation of males is higher than females** but still the difference in predicted values is not that big. While comparing the *males* to the females who *never married* , We found out that the *never married women* mean is less than the *married women* mean but also the difference is small.

We’ve got a mean of 0.8639 for married-spouse present women while a mean of 0.7913 for married-spouse absent women . compared to a mean is 0.92 for married male .

So the men participation in the labor force gets the highest fraction , followed by the married-spouse present , and then the married-spouse absent , and by comparing this with the never married women , we’ve got a mean of 0.8157 which is pretty close to all of them but still higher than the married-spouse absent mean . So that might ensure the high participation of the married women in the labor force , and that was also indicated before using the two Figures above that compares between the married and the never married women in the labor force .

In general , These statistics so far indicate that all the independent variables are exogenous except for some or few of their categories while the education variables get almost the highest estimate which indicates that Education is one factor that affects the labor force participation.

**For the NHIS \_2021 data :**

We’ve done the logit model between the dependent variable and the independent variables that mentioned above. The model came out as a significant model to be in the workloss except for the RECVACC (vaccination receiving ) variable with only no vaccination receiving . as it barely affected the Workloss.

And this will be the first logit model formula :

model\_logit1 <- glm(WRKLOSSRV ~ WORRY + HADCOVID + ANXIOUS + RECVDVACC + KINDWORK + GENID\_DESCRIBE , family = binomial, data = data\_new)

These were the primary results of regression besides that as we mentioned before the model came out significant with a low p-value less than the significance level .

Call:

glm(formula = WRKLOSSRV ~ WORRY + HADCOVID + ANXIOUS + RECVDVACC +

KINDWORK + GENID\_DESCRIBE, family = binomial, data = data\_new)

Residuals:

Min 1Q Median 3Q Max

-4.1998 0.0212 0.0706 0.0813 2.8849

After running another logit model by adding the (works\_remote) variable , this variable significantly affects the model with a 0.1 significance level , So in these two logit models , all the independent variables are positively related to the workloss variable .

So regression equation for model logit 1 will be as following :

WRKLOSS = -3.73073 + 1.5592 WORRY + 1.5206 HADCOVID + 2.5764 ANXIOUS + 0.2879 RECVACCdid not get vaxx + 1.80843 KINDWORK + 0.30363 GENID\_DESCRIBEmale.

And regression equation for model 2 will be as following :

WRKLOSS = - 3.8877 + 0.3230 GENID\_DESCRIBE + 1.4734 HADCOVID + 2.5507 ANXIOUS + 1.5721 WORRY + 1.8586 KINDWORK + 0.2533 works\_remote + 0.3037 RECVACC did not get vaxx .

So All of the independent variables are good predictors so far , except for the RECVACC variable with no vaccine receiving which rarely affects the workloss in 2021 which I see is contradicted with reality as not having a vaccine was a big issue in the workplace especially at this period of time .

About the Kind of work Variable , it's one of the independent variables that strongly affected on the workloss and this is ensured with what Stefania Albanesi and Jiyeon Kim found in their paper as the types of occupations played a very big role in losing the work and affect the rate of participation in the labor force , while Stefania Albanesi and Jiyeon Kim used a different data and their paper was in 2019 at the very beginning of the COVID-19 pandemic , and the data that we used in this paper is the NHIS data which was in 2021 (one and half year after COVID start).

By adding Education variable to the model , We ran another logit model :

model\_logit3 <- glm(WRKLOSSRV ~ GENID\_DESCRIBE + HADCOVID + ANXIOUS + WORRY + KINDWORK + EEDUC + works\_remote + RECVDVACC, family = binomial, data = data\_new)

From this regression model results , we found out that the education does not statistically affect on the work loss model which is something contradicted with all the research papers that we’ve studied and mentioned that the education plays a very strong role in the labor force participation and also this contradict with our other data analysis that we did before on the ACS\_2017.

Another thing that we were supposed to mention since we started our first model on the NHIS data is that also the WORRY variable does not have a big effect on the work loss while being ANXIOUS has a big effect on the work loss , which is something that is confusing as they both almost have the same meaning .

And by doing another model adding the Race variable to it , we didn’t find it that useful to add it to the model as it also does not affect the work loss that much . but in general our model is still significant by most of its variables.

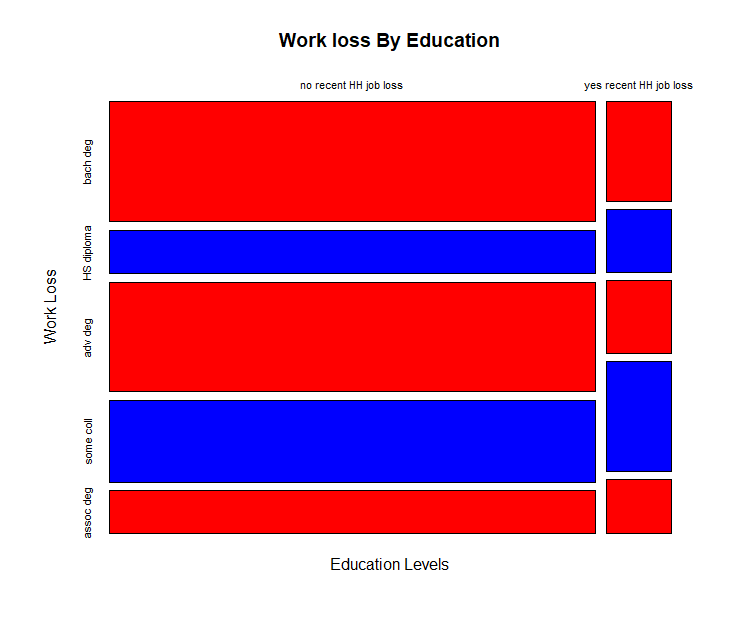
By changing the subset group this time and doing it as only people who are 22 years old and over, and by running a simple logit model:

model\_logit5 <- glm(WRKLOSSRV ~ EEDUC + MS + RRACE + GENID\_DESCRIBE,

family = binomial, data = dat\_use1)

We’ve got that the model is still significant except for the gender described .

By doing a graph indicating the work loss participation by Education levels , we’ve got this graph below which indicates the big work loss is coming from the bachelor degree recipients and those with advanced degree.



After that we did a probit model which we got a statistically significant model :

model\_probit1 <- glm(WRKLOSSRV ~ EEDUC + MS + RRACE + GENID\_DESCRIBE,

family = binomial (link = 'probit'), data = dat\_use1)

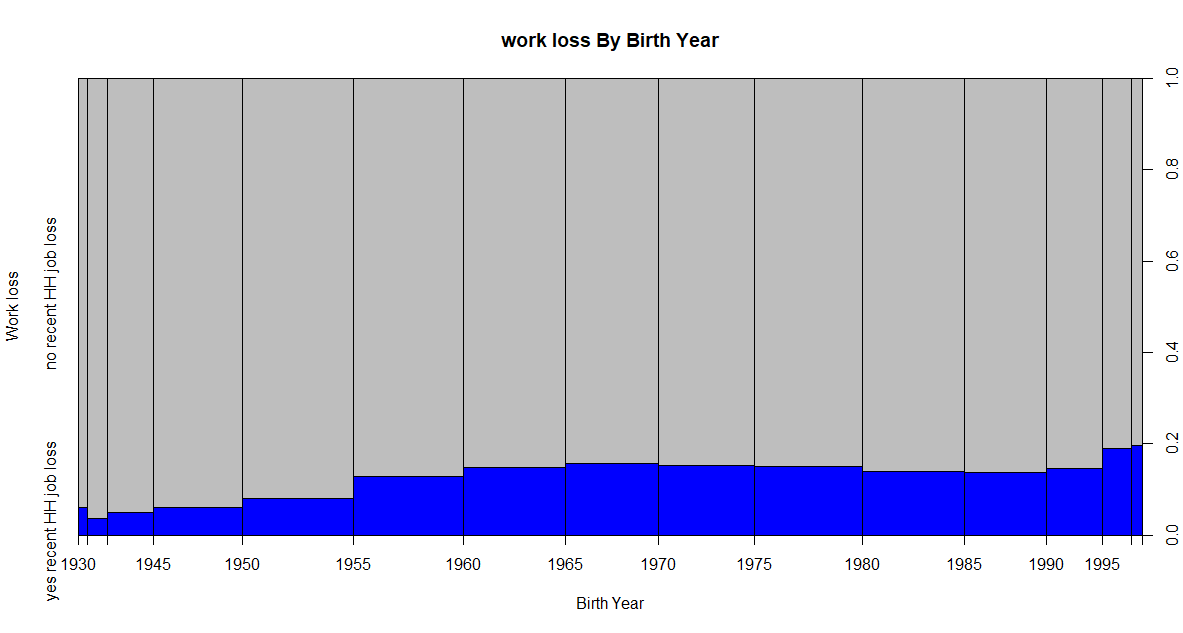
And by adding the Income variable in the model , we’ve got the same result as before from the simple table stats , that most of groups who affected on the workloss who are those with HH income less than $25k and $50k - 74.9 while the rest don’t affect on the work loss variable at all .

After that we added some polynomials to the model by cubing the income variable but the results were still the same .

By comparing logit 5 by the probit 1 :

They both give strongly significant models with results while their variable estimated values are different and it looks a little bit bigger in favor of the logit model , but still both can work while checking the work loss. And also still in both models the "GENIDESCRIBEmale" is the only insignificant variable. And the accuracy level of the model is 91% .

By doing a graph of workloss by Birth Year , We see people who got born the "2000's" are the most people who lost their jobs .



And by using proportions , we see only %11 of the sample recently lost their jobs compared with % 88 who haven't lost their jobs recently ! .

no recent HH job loss yes recent HH job loss

0.8800054 0.1199946

And while doing the summary for the reason for work loss , being "sick or disabled" showed a big number of people after being "retired"

And the "laid off" results takes the top number of people followed by " retired" and then "am/was sick w covid or caring for sick w covid" comes after.

We found a challenge doing the more complicated regressions in the NHIS\_2021 data as we could not find a result at the end.

**The Confusion Matrix :**

**For the ACS\_data 2017 :**

By doing the confusion matrix for those categories of race , marital status and some of educational categories who are not participating in the labor force in our model, There are (4034) Type I errors compared to (287) Type II errors , So since type II is false-negative , The model should be good as type II errors are less and the independent variables can be good predictors .

**0 1**

FALSE 344  **4034**

TRUE **287**  18663

**Sobj model :**

**For the ACS\_data 2017 :**

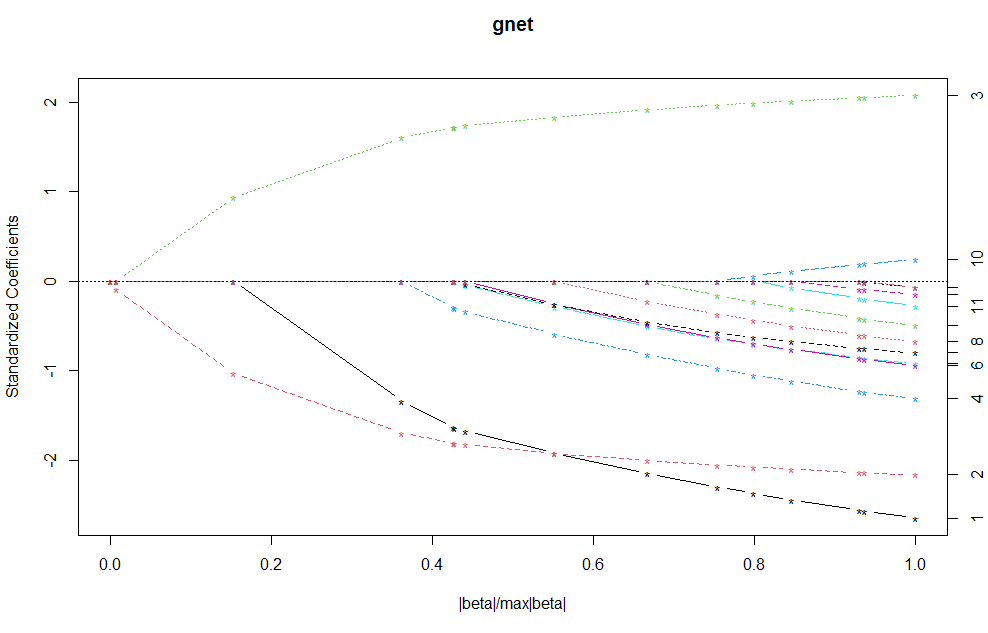
We created a data frame and model matrix afterwards and did a Sobj model and while doing a linear model according to that , we found the model has some variables which are not that significant then but still the model as a whole significant with the same p-value: < 2.2e-16 where we can reject the null hypothesis of no relation . and again the Type II errors are less than Type I errors , and while doing the logit model , we found out that the logit model is better than the linear model in predicting as its accuracy level is 0.67 compared to 0.36 for the linear.

**More Complicated regressions :**

**For the ACS\_2017 data :**

We did more complicated regressions , using the random forest method , Spike and Slab, Elastic Net , and LOESS in the R file .

For the Spike and Slab method , we plotted a graph of the standardized coefficients which will looked like this :



**Conclusion :**

In this paper we were trying to figure out the behavior of labor force participation , and how it differs along with different factors , which could be demographic or something external such as illness or disability or shock like the Covid Pandemic .

We were trying to combine this behavior accordingly with the Covid pandemic using both the *ACS* and *NHIS* data and shows the decline in the labor force for both males and females nowadays due to different factors.

***The most interesting result***  is that the ***married females***contribute more in employment than the***unmarried ones****,* who are most likely to have more spare time and don’t have as many responsibilities as the married ones.

Not all the times adding **polynomial variables** can affect the model as it depends on what we are analyzing and depends on the rest of the independent variables too.

The loss of women employment because of the pandemic may leave a long term effect in the labor force despite the beginning of recovery unlike those married women who became mothers who used to leave their jobs TEMPORARILY to look after their children and go back to their jobs which was before the pandemic , and that may still cause losses in earnings and wages in the economy which also may lead to a slow progression for the economy.

Another effect that has been arisen after the COVID-19 pandemic was the remote work which expected to be lasted after the pandemic period finishes and will be an essential factor in the labor force environment which would help too many segments of workers who left their jobs before because of the fear of infections and interactions , and this will help a lot in the economy recovery and help and encourage a high percentage of women to participate in the labor force more specially in these services providing and managerial jobs.

**Bibliography :**

The Changing Cyclicality of Labor Force Participation [ By By Willem Van Zandweghe]

<https://www.kansascityfed.org/Economic%20Review/documents/589/2017-The%20Changing%20Cyclicality%20of%20Labor%20Force%20Participation.pdf>

Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate.

[pp. 1-59 , Published by: [The Johns Hopkins University Pres](https://www-jstor-org.ccny-proxy1.libr.ccny.cuny.edu/publisher/jhup)s]

Labor Force Statistics from the Current Population Survey [https://www.bls.gov/cps/home.htm]

The Effect of Women's Labor Force Participation on the Distribution of Income in the United States.

[Author(s): Judith Treas Source: Annual Review of Sociology , 1987, Vol. 13 (1987),

pp. 259-288 Published by: Annual Reviews Stable].

URL: <https://www.jstor.org/stable/2083249>

WHAT WE KNOW AND DON’T KNOW ABOUT DECLINING LABOR FORCE PARTICIPATION: A REVIEW [Eleanor Krause and Isabel Sawhill The Brookings Institution May 2017]

Where Is Everybody? The Shrinking Labor Force Participation Rate .

[BY MICHAEL DOTSEY, SHIGERU FUJITA, AND LEENA RUDANKO]

<https://www.philadelphiaf>

ed.org/-/media/frbp/assets/economy/articles/economic-insights/2017/q4/eiq4\_where-is-everybody.pdf

# The evolving impacts of the COVID-19 pandemic on gender inequality in the US labor market. [[Kenneth A. Couch](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Couch%2C+Kenneth+A),[Robert W. Fairlie](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Fairlie%2C+Robert+W),[Huanan Xu](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Xu%2C+Huanan)]

<https://onlinelibrary.wiley.com/doi/full/10.1111/ecin.13054>

The Changing Cyclicality of Labor Force Participation <https://www.kansascityfed.org/Economic%20Review/documents/589/2017-The%20Changing%20Cyclicality%20of%20Labor%20Force%20Participation.pdf>

Effects of the COVID-19 Recession on the US Labor Market: Occupation, Family, and Gender [Stefania Albanesi and Jiyeon Kim].

<https://www.jstor.org/stable/pdf/27041212.pdf?refreqid=excelsior%3A46cb9350f89a0f0967c7d503caa6be92&ab_segments=0%2FSYC-6646_basic_search%2Fcontrol&origin=&acceptTC=1>

# U.S. labor market inches back from the COVID-19 shock, but recovery is far from complete [[RAKESH KOCHHAR](https://www.pewresearch.org/staff/rakesh-kochhar) AND [JESSE BENNETT](https://www.pewresearch.org/staff/jesse-bennett)]

<https://www.pewresearch.org/fact-tank/2021/04/14/u-s-labor-market-inches-back-from-the-covid-19-shock-but-recovery-is-far-from-complete/>