The final essay to the test

Dian Zhang, 2016110592

School of Public Economics and Administration

The purpose of this essay is to demonstrate the answer to the Admission Test to "The 2nd Empirical Research Immersion" at SUFE Finance. This essay consists of two parts, corresponding two exercises.

◆ Part One

The first exercise is just a test on data processing. The data set is about American Medicare Advantage (MA) enrollment by contract-county pair. Because the file is in format of csv, it should be turned to .dta. Then, turning core column such as eligible, enrollee and penetration from strings to numbers.

The beginning step should be data observation. The question describes the dataset as "a contract-county pair", which means contract and county may be a double pointer. Looking at those columns describing columns, countyssa exclusively refers to every county, which has two attributes: county name and the state it affiliated to. Meanwhile, contract also exactly represents various Medicare plan, whose name and type are unique.

Column of state lists 50 states, 1 WDC, 5 island commonwealths and 2 special values. The latter two should be deleted in this work.

The final submission of the title is a county-level data set (i.e., a county-by-row, giving a descriptive statistical data set) with new additional information:

Numberofplans1: number of health plans with more than 10 enrollees
Numberofplans2: number of health plans with penetration > 0.5

Therefore, the calculation should be based on the classification of counties. New variables will be used to assist finishing this job.

N1 identifies the number of entries (ENR) greater than 10 (bool).
N2 identifies items with a penetration rate of pen greater than 0.5 (bool).

Then, summing up the value of each county's bool value. Finally, delete the auxiliary variables N1 and N2.

Another two newly added variables are total enrollees ans total penetration. Just sorting data by county and then use "total" function.

◆ Part Two

Notes: At the end of this article, there is a summary of the second part. The reader can read the final summary first and follow me through the whole research process from the beginning.

In the second exercise, a dataset on labor force is given to analysis. There is one main problem:

How have hourly wages ("wage") and labor force participation ("lfp") evolved for skilled and unskilled workers since 1976?

with 4 questions:

a) Please summarize the key trends for wages and labor force participation. (Basic)
b) Among men older than age 25, which groups of people have had the biggest changes in labor force participation? (Basic)
c) What factors do you think are driving these patterns? (Intermediate)
d) What evidence might you want to assemble to test these hypotheses if you were to investigate them further? (Difficult)

What I have done in this paper:

- Exploring the labor participation rate and hourly salary of different groups according to different age, sex, race and education level.
- Expounding why education is the most important factor and exploring the relationship between education and other factors.
- A preliminary analysis of the possibility that education is influenced by other factors.

First, let's watch the dataset. We can see that this is a sample like spot checks separately recording information of samples, such as age, sex, occupation, race, education, labor income status, etc.

Among them, **Age** has groups. **Ethnic** groups include large groups, Latino groups, white bool variables. **Education** has specific educational status and whether it has received higher education. **Labor** has the Employment Status ID, the state of labor, the number and grouping of working weeks last year, the average working hours per week, etc. **Income state** includes total individual income, individual wage income and working hour income.

● Start

It's time to solve the problems.

✧ Graph 2-1 illustrates the trend of participation rate of all labor force since 1976.

Graph 2-1

The picture shows that before 2001, the participation rate of labor rate is climbing. However, after 2001 reaching its peak, the rate starts to decrease. Here comes a question: **Why labor participation rate has declined in this century?**

In addition, in subsequent studies, I found that **there seemed to be a "big change" in the data around 2000.** Therefore, when thinking, I may tend to observe the differences between the centuries.

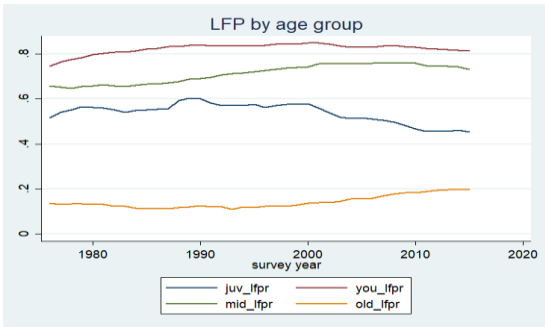
✧ Graph 2- 2shows the trend of hourly wage of all labor force since 1976.

Graph 2-2

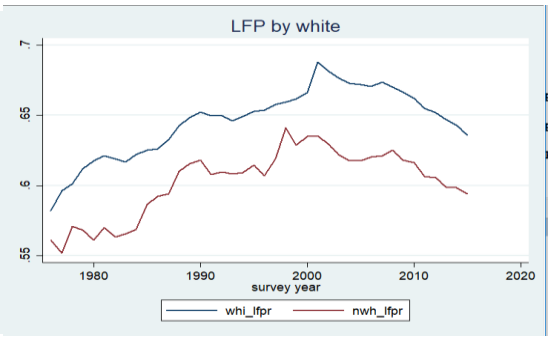
Hourly wages are rising steadily.

✓ **Next, dividing people into different groups to further explore the law of labor participation rate.**

Graph 2-3 divides people by age group. Graph 2-4 by race (white).

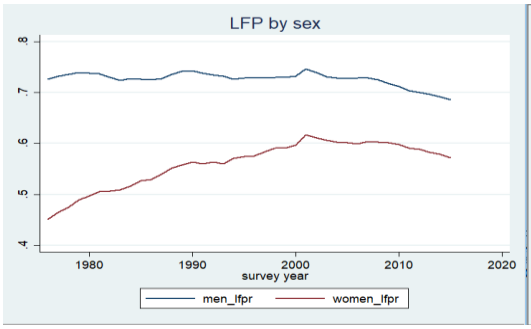


Graph 2-3

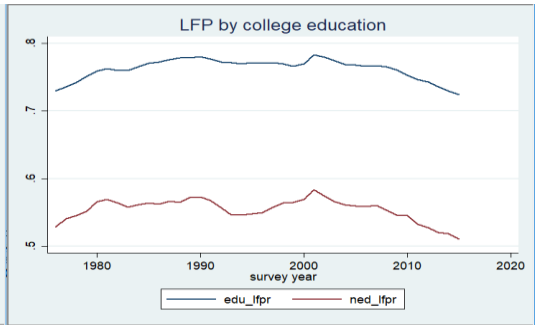


Graph 2-4

Graph 2-5 by sex. Graph 2-6 by college education



Graph 2-5



Graph 2-6

In Graph 2-3, juvenilia refer to people younger than 25 years old, young people refer to people 25-45 years old, middle-aged people refer to people 45-65 years old, older people refer to people older than 65 years old.

Obviously, the biggest gap occurs between the elderly group and the youth group. The most significant change comes from a marked increase in the female labor force participation rate.

But, astonishly, the disparity in labor participation rates caused by education has not changed much in the past few decades.

From the above chart we can draw: from the age group, except for the labor participation rate of elderly group being increasing, other groups have decreased to varying degrees. Juvenile group has the largest withdrawal.

The gap between men and women has been shrinking dramatically in the last century, but this phenomenon has stalled since the beginning of this century, and the overall participation rate has shown a downward trend.

The difference in labor participation rates between the groups with and without higher education remained basically unchanged. In the uneducated group, the labor participation rate curve is not smooth as the educated group, reflecting to some extent that the uneducated are more likely to be unemployed for a long time or to launch the

labor market in the recession.

As for race, although the disparities between races are not as big as the news and our imagination, they seem to have reversed the trend of decreasing disparities between races since the beginning of this century.

Based on the above data, several possible reasons can explain the decreasing of labor force participation:

- Overall labor force unemployment and exit
- Ageing of population.
- The decline of white population in population structure.

a) Potential reason: Overall labor force unemployment and exit

Figure 2-7 shows the data on the proportion of wages to total income since 1976.

Graph 2-7

The proportion of wages to total income declined in 1980-1990 and this century, and this proportion only recovered in the last decade of the last century, reaching the level of the early 1980s. This rate may in some sense explain the long-term unemployment or withdrawal from the labor market, resulting in a reduction in wages. People are bound to need a certain source of income for their livelihood. **But this logic does not seem to be able to explain what happened in the last centu**ry.

b) Potential reason: Ageing of population

Figure 2-8 shows the demographic changes in the sample data.

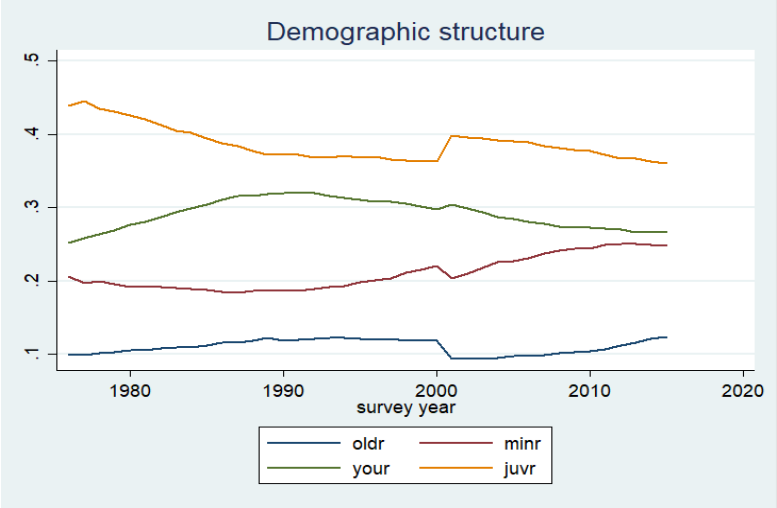


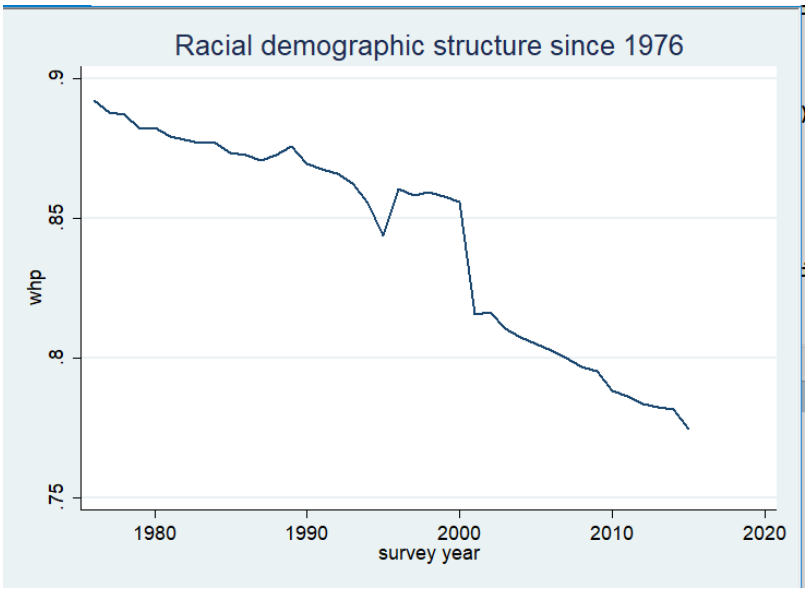
Figure 2-8

As you can see from the picture, the population in the sample data is aging. The proportion of juvenile and young people is declining, and the proportion of middle-aged and elderly people is rising. Ignoring a simultaneous fluctuation in the beginning of new century.

Considering the labor participation rate (Graph 2-3), the youth group with the highest labor participation rate is declining, and the middle-aged and elderly groups with the lowest labor participation rate are increasing. Such data seem to partly explain the decline in the labor force participation rate of the sample data in this century.

c) Potential reason: Changes in racial demographic structure

Chart 2-9 shows the change in racial demographic structure.



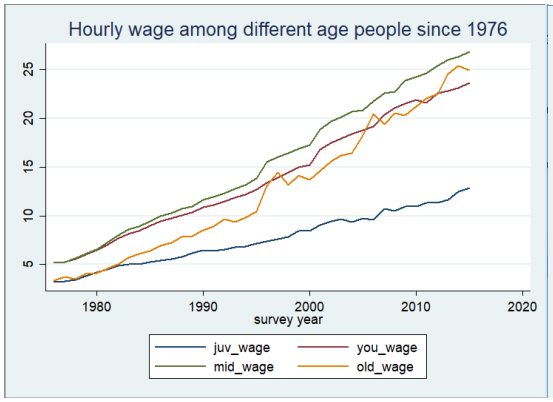
Graph 2-9

Figure “whp” is white population proportion. We can see from the chart that the proportion of white people in the statistical sample has been declining over the past 40 years. Similar to almost all of the data, something unknown happened in 2000 that caused the data to change dramatically (guess factors such as changes in statistical caliber). Thus, look at the data of this century and the last century. In the last century, the white population did not seem to have fallen as fast as this century.

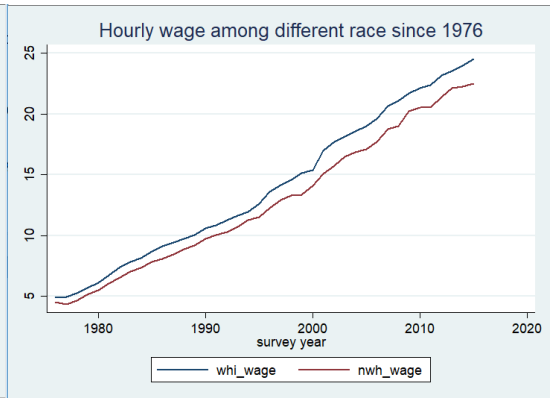
The accelerated decline in the proportion of white population may be a potential secondary reason for the change in the labor force participation rate.

Note: the white population data seem to be slightly different from that of the US Census in 2010. The proportion of whites in census data is about 60-65, while the proportion in the sample data seems to be above 70. Unfortunately, I don't have enough time to explore “The numbers game”.

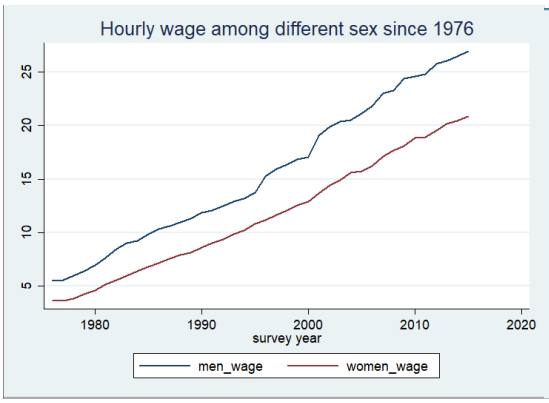
- **Then, let’s look at hourly wage.**



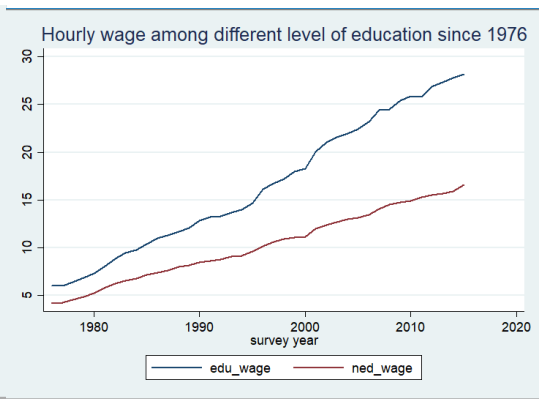
Graph 2-10



Graph 2-11



Graph 2-12



Graph 2-13

As you can see from the chart above, middle-aged people have the highest income and young people the lowest. Apart from young people, the income gap of other groups is not large. Like labor participation, racial differences are very small. The wage gap between the sexes is widening at a slower pace, which contrasts with the rate of labor participation. The pay gap between the educated and the uneducated presents a

shocking tear over time.

From the perspective of income level, the biggest gap and change come from education level.

● Without Juvenile

The above data processing is based on the whole people. Next, we remove juvenile and look at the rest.

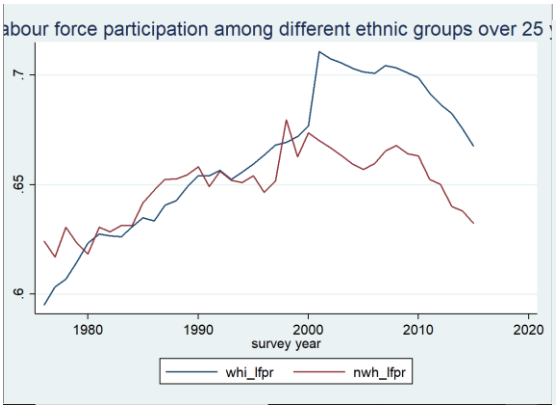
✓ Overall labor force participation rate and wage trend

Graph 2-14

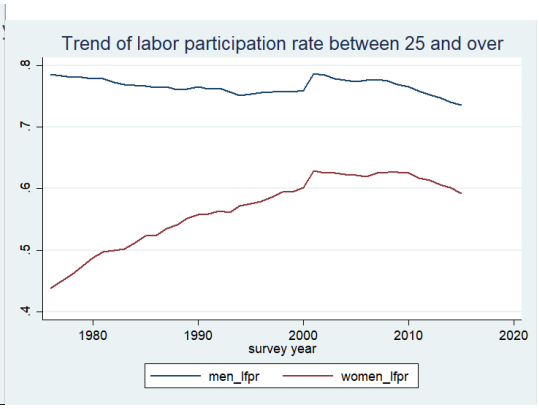
Graph 2-15

The overall trend does not seem to be very different from the overall trend before.

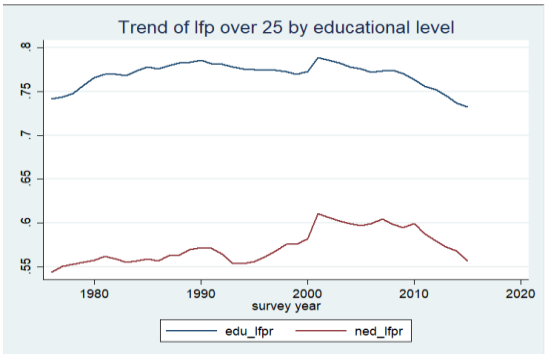
✓ Trend of group labor participation rate.



Graph 2-16



Graph 2-17



Graph 2-18

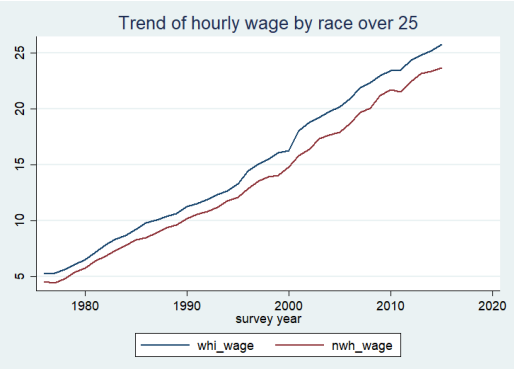
The other two elements are not very different. The only difference comes from race. As we can see in the previous article, there has always been a disparity in labor participation rates between races, but the disparity remains largely constant. But now this picture tells us another story. In the last century, there was no difference between races. By the new century, however, the labor participation rate among ethnic groups has shown a great gap.

Taking into account what has been said before, there is a strange change in the sample data around 2000. If we ignore the changes brought about by this period and refer the data of this century to those of the last century in a certain way in our minds, we can conclude that we can find that the data of different ethnic groups' labor participation rate are changing in the same direction.

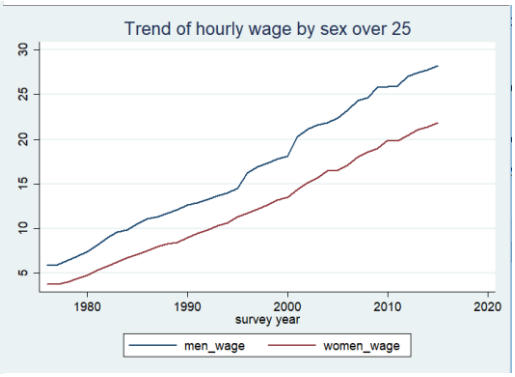
This strange phenomenon also seems to tell us that "The Big Change" in this statistic at the turn of the century may have been due to the change of white workers over 25.

I don't have time to go into the details of why this is such a strange phenomenon. However, it can be inferred from the known information that this may be due to the fact that a certain US statistical survey department has adopted a different statistical caliber and indicators in the new century.

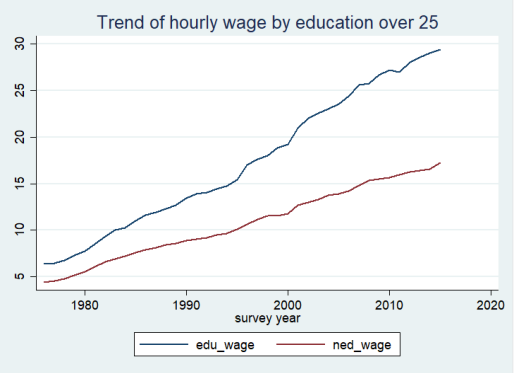
✓ Group salary trend.



Graph 2-19



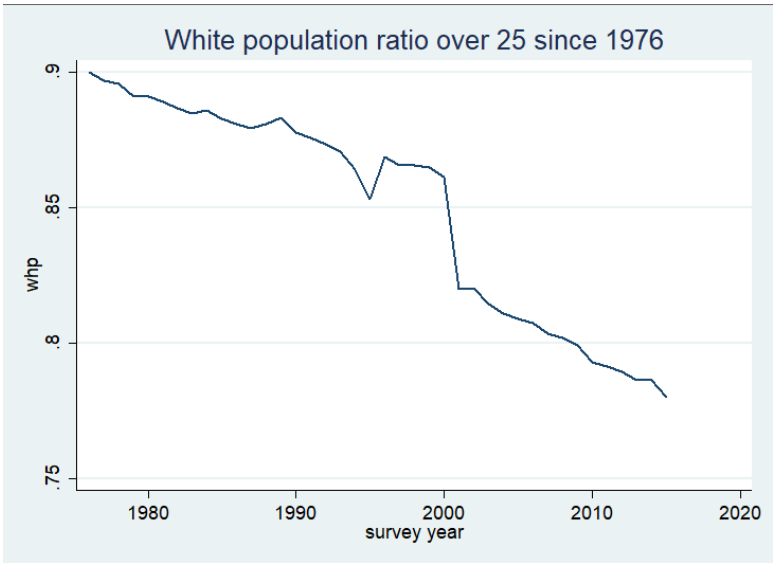
Graph 2-20



Graph 2-21

These are not very different from the previous data.

✓ White population structure



Graph 2-22

It doesn't seem to make much difference.

Observing the labor data without juvenile, we can draw the following conclusions: The biggest wage gap is still due to different levels of education, and the biggest gap narrowing is due to women's labor participation rate.

● Analysis

Let's first analyze the differences in educational level. **Whether or not having higher education has brought very significant difference to the future income level.** Or, in other words, education is, to a certain extent, the most important factor. This may also partly explain the disparities we find in the age group: young people's incomes are far lower than those of their predecessors. Generally speaking, younger group should have the largest proportion of those who have not received higher education (because they are

not yet old enough to receive higher education).

Why does education become more and more important? Let's think carefully about the changes brought about by science and technology over the past forty years. These forty years are the forty years of rapid development of computer and Internet technology. Because of the progress of science and technology, many low-tech jobs which originally needed manpower were eliminated, and many high-tech research and development work were born. A recent obvious example is auto-driving technology. The rapid development of automatic driving technology will make society less and less need of drivers, who are often without high-education. On the contrary, the demand for program developers for auto-driving systems continues to grow. However, training a high-tech person in society will not be faster than science changing the world. As a result, insufficient supply of high-skilled personnel leads to higher salaries, and lower demand for low-skilled jobs leads to lower relative salaries.

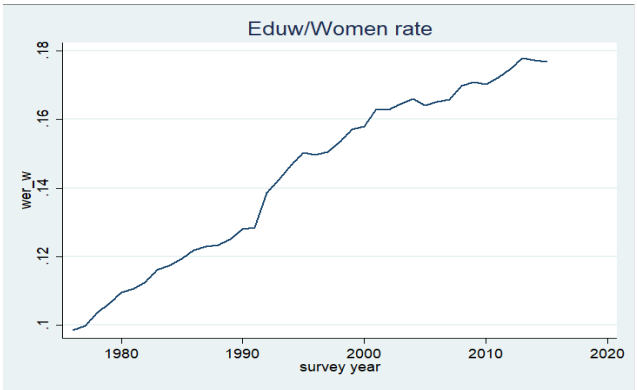
Therefore, we can put forward a **hypothesis that technological progress has led to such a wage gap.** If we're going to explore this conjecture further, maybe I need a copy of the salary at graduation and the average five-year salary at college in the United States. If this conjecture is correct, then even among highly educated people, the employment rate, graduation salary and salary growth of science and engineering students should be much higher than that of history and literature students.

Another possibility that will impact people with low educational level is globalization. The past 40 years have been marked by an increasing degree of globalization, which has led to the flow of elements and more adequate competition. According to Ricardo's theory of comparative advantage, in the United States, the demand for low-skilled workers in labor-intensive industries will be reduced because the leading capital and technology in the United States can be globalized to acquire labor production factors around the world. A vivid example of life is that the United States is now importing a large number of clothing, light industrial products from China, Vietnam and other countries, cheaper and more skilled workers than the United States produced. Under the severe competition of globalization, the interests of the disadvantaged ordinary workers in the United States have been hurt.

Now let's take a look at the relationship between education and gender and race.

About gender differences. **The hypothesis I would like to make here is that the Second Women's Liberation Movement**, launched in the 1960s, has greatly enhanced the social status of women. As a result, women are beginning to have more educational opportunities and thus better skills and abilities to survive in the workplace. Excellent performance also brings the whole society to give women a higher evaluation, thus recognizing women, forming a virtuous circle. As a result, the difference of labor participation rate between men and women has been rapidly reduced in the last century.

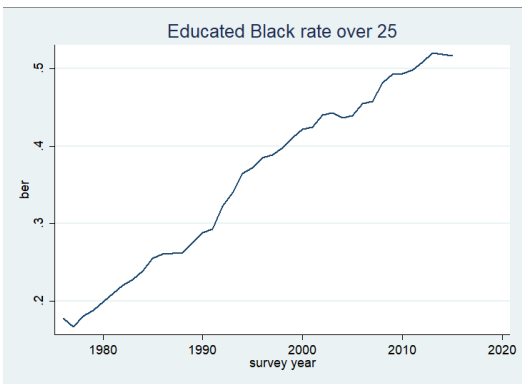
As for the stagnation of the gap, it may be that women's position in the workplace is difficult to improve further. Because of women's reproductive responsibilities and costs, some women who choose to raise their offspring launch the labor market and become "full-time mothers".



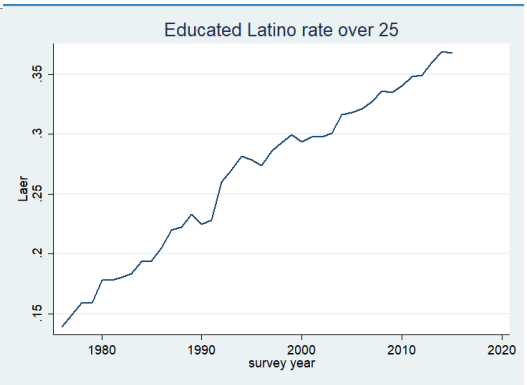
Graph 2-23

The above picture shows the proportion of women who have been educated in higher education over 25 since 1976. We can see that the proportion of highly educated women in the female population is rising, and partly confirm previous speculations about feminism.

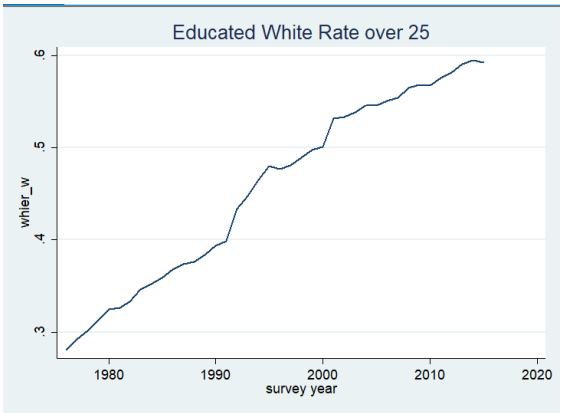
With regard to race, this is a very complicated and thorny issue. Let's look at the educational level of several ethnic groups as well as women.



Graph 2-24

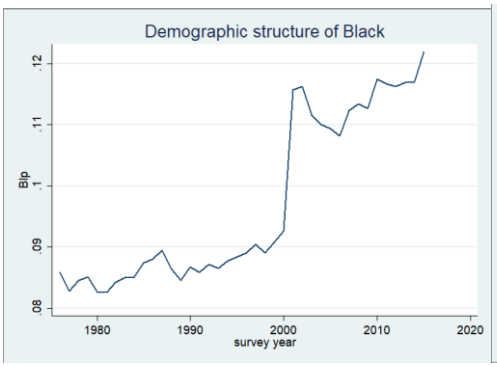


Graph 2-25

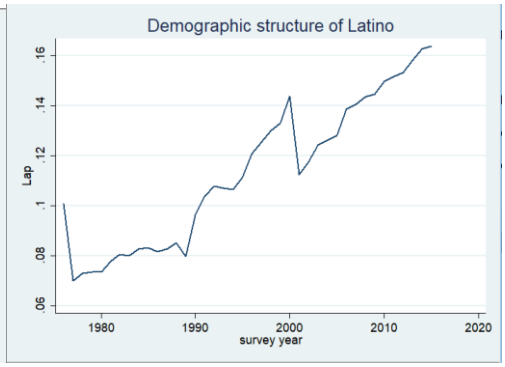


Graph 2-26

The above three pictures reveal the educational level of the three main ethnic groups over 25 years of age. Despite the dramatic rise in educational attainment across all ethnic groups, there is a huge difference in absolute values: whites are the most educated, blacks are the least educated, and Latinos are far behind.



Graph 2-27



Graph 2-28

The above data also reflect that the minority population in the sample is generally increasing. (Let's ignore the strange phenomena at the turn of the century.) Minorities with lower education levels are on the rise, reducing their wages and labor participation rates. However, because minorities account for a small proportion of the statistical sample, there is no significant inter-ethnic disparity. The result seems to have provided some details.

Let's talk about something else. To some extent, the significant increase in white data corresponds to a significant decline in other ethnic data. **This may involve some problems on illegal immigrants.** Due to time constraints, I do not have the opportunity to investigate the situation of illegal immigrants in the United States in detail, but I estimate that there are more and more illegal immigrants in the United States. After all, we never heard much about illegal immigration in the news before America was so strong. Now, illegal immigration seems to be a very big social problem in the United States.

Illegal immigrants are mainly ethnic minorities. The large number of illegal immigrants would make it impossible for the authorities to investigate in detail the labor situation of some ethnic minorities. This will reduce the labor force participation rate and wage level of ethnic minorities.

The proportion of blacks and Latinos in the 2010 U.S. census is roughly the same as this sample. **But there seems to be some discrepancy in the proportion of whites.** The proportion of whites in census data is about 60-65, while the proportion in the sample data seems to be above 70.

In the absence of time to investigate the background of data, it is reasonable to doubt the quality of data. **Another skepticism about the quality of race related data comes from official motives.** As we all know, racial equality has become "politically correct" in the United States. And the United States is a democratic country, both Democrats and Republicans will want to show some good indicators when they take office. Even if the central government has no such motive, the local government is still a small "democratic mechanism" with the same reason for reasonable doubt.

When it comes to logic.

This data is actually a relationship between labor conditions and workers' gender, race, age and education. In previous data collation, we've seen in previous data collation that education seems to have the greatest correlation with the overall labor situation. On this basis, the relationship among race, sex and education level was further observed with the education level of workers as the core. Then, what is the further? The thinking before was that a certain factor (such as education) determines the level of labor income. If we think about reverse logic, does the level of labor income affect the level of education?

This may be to investigate the relationship between the income level of parents and the education level of their children. If the income level of the parents leads to a higher level of education of their children, a vicious circle arises from the gap between the rich and the poor.

Conclusion

Having written so much, it's time to make a conclusion.

What I have done in this paper:

- Exploring the labor participation rate and hourly salary of different groups according to different age, sex, race and education level.
- Expounding why education is the most important factor and exploring the relationship between education and other factors.
- A preliminary analysis of the possibility that education is influenced by other factors.

What I still need to do in the future:

- To find out the source of data and the specific criteria of each subject.
- To explore the impact of structural changes in labor management demand on wage level caused by technological progress. A copy of the salary at graduation and the average five-year salary after graduation in the United States may be needed.
- To explore the impact of globalization on the labor situation of American workers
- To investigate the situation of illegal immigrants in the United States in detail.
- To explore the impact of parental income and other potential factors on the education level of offspring.