

# Examination for the Decrease in US Birth Rate After the Great Recession: the Change of Hispanic Fertility Rate and Education Convergence Behind It

Zelong Yu, Yiyi Ren, Zihan Jin

02/22/2022

## Introduction

## Data

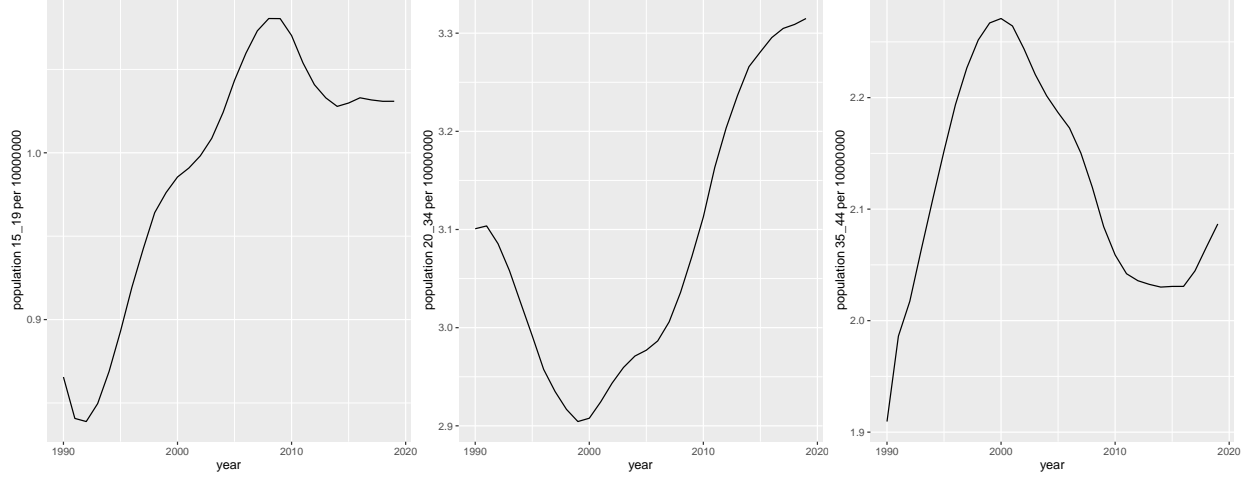
The basic data all comes from the Centers for Disease Control. The datasets provided by the original paper writers are also beneficial to this paper. I analyzed all these using R package including: Rreadr(@), tidyverse(@), ggplot2(@), haven(@), gcookbook(@), ggrepel(@), gtrendsR(@), usmap(@), dplyr(@), car(@), patchwork(@).

There are several datasets we used in this project. The first dataset is “figs\_1\_2\_3.csv” which was provided by Kearney, Melissa S., Phillip B. Levine, and Luke Pardue, the writer of the original paper. The original paper is named “The Puzzle of Falling US Birth Rates since the Great Recession”. This paper mainly argues the US birth rates of Hispanic fertility dropped significantly over the years after the Great Recession. The authors studied a lot of factors including age, education, race and ethnicity, marital status, and birth parity. Based on these, they find out the downward trend is society-wide also rarely reversible in the next few years. The authors also produce another dataset called “figs\_2a\_2b.csv”, which is useful in our replication project as well. The two datasets “figs\_1\_2\_3.csv” and “figs\_2a\_2b.csv” contain many columns which include year, different types of birth rate, and descriptions of women’s characters. Among 33 variables, there is only half of them that are meaningful to our replication interests, hence 10 variables from “figs\_1\_2\_3.csv” and all 10 variables from “figs\_2a\_2b.csv” are used in our project.

“age\_race\_comp\_seer” is another dataset that contains lots of informative variables of the total population. It is produced by the Centers for Disease Control and Prevention Surveillance, Epidemiology, and End Result (CDC SEER). CDC is founded in 1946 and it is an agency of the U.S. Department of Health and Human Services that provides reliable information to protect public health and safety (<https://seer.cancer.gov/about/overview.html>). CDC collected the data through the National Notifiable Diseases Surveillance System (NNDSS). The dataset we used in this project is also collected by about 3,000 health departments from local cities. The dataset “age\_race\_comp\_seer”, separates the data by different years and states. This dataset contains feathered variables of age, race, and the combination of age and race. The variable of age offers the information of the exact population of US women by 5 different age groups, which are age 15-19, 20-24, 25-34, 35-39, and 40-44. Besides, there is a group of variables in the data that give the information of population for each race as well as for a specific age group. For example, “popwhitenh1519” gives the total population of white women aged between 15-19.

“nchs\_births\_pop\_1990\_2019.dta” is delivered by The National Center for Health Statistics (NCHS). NCHS is a department under control by Centers for Disease Control and Prevention Surveillance (CDC), which provides statistical information to support CDC to prevent health crises from happening. NCHS collects the data using administrative records and medical records from the related facility. In this dataset,

it presents 264 variables, grouped by name and state. The main data contain information related to nationality, ethnicity, education level, and population. Among them, “numbirth1544” is the variable we aim to analyze later. “numbirth1544” means the number of births given by the mother whose age is between 15-44 years old.



## Reproduction

- Figure 2, Panel A and B:

We adopted the dataset “figs\_2a\_2b” provided by the authors to reproduce panel A and B in figure 2, which named as “Five-year Age Group from 15 to 44” and “Race and Ethnicity from 15 to 44”. To begin with, we renamed the variables into “year”, “15-19”, “20-24”, “25-29”, “30-34”, “35-39”, “40-44”, “White\_non\_H”, “Black\_non\_H”, “Hispanic”. Then, we collected a set of subgroups together using function gather to form a new dataframe and plotted the ggplot. In panel A, we gathered 6 age groups, including “15-19”, “20-24”, “25-29”, “30-34”, “35-39”, and “40-44” from the dataset “figs\_2a\_2b”. The rest variables in the dataset, including “White\_non\_H”, “Black\_non\_H”, and “Hispanic”, are selected to a newly created dataframe for panel B. Finally, with dataframe ready for each panel, we draw their graph using ggplot.

- Figure 2, Panel C, D, and E:

We used the dataset “figs\_1\_2\_3” provided by the authors to reproduce panel C, D, and E in figure 2, which named as “Hispanic subpopulation from 15 to 44”, “Mother’s level of education from 20 to 44”, and “Marital status from 15 to 44”. Firstly, we located the related four columns by index 20, 21, 22, 23 and renamed them as “Native\_born\_Mexican”, “Foreign\_born\_Mexican”, “Native\_born\_non\_Mexican”, “Foreign\_born\_non\_Mexican” for panel C; located columns by index 13, 14, 15, 16 and renamed them as “No\_high\_school\_degree”, “high\_school\_degree”, “Some\_college”, “College\_graduate” for panel D; and located columns by index 17 and 18 and renamed them as “Unmarried” and “Married” for panel E. Then, we filtered out the missing value and selected columns by year. We gathered the value and rename the three columns as “year”, “Hispanic”, and “value”. At the end, we created graphs for each panel by ggplot.

- Figure 2, Panel F:

We used the datasets “age\_race\_comp\_seer” and “nchs\_births\_pop\_1990\_2019” provided by the authors to reproduce panel F in figure 2, which named as “Parity from ages 15 to 44”. First, we merged these two datasets together by “sname” (state name) and year. Then, we added up a new column called “pop2044”, which is created by calculating the sum of “pop2024.x”, “pop2534.x” and “pop3544.x” using function mutate.

- Figure 3:

Figure 3 is a state-level US map that indicates the changes in birth rates among women aged 15 to 44 between two five-year periods (2004-2008, 2015-2019) before and after the Great Recession. We plotted this figure using the datasets “nchs\_births\_pop\_1990\_2019” and “numbirth1544” provided by the authors. To begin with, we filter out the missing value and observations outside the two five year period. We dropped the states with names of “AB”, “BC”, “MB”, “NB”, “NS”, “ON”, “QC”, “SK”, “XX” as well.

There are many columns in each dataset, so I selected some useful parts of them to found new dataframe after I filtered. For “state\_births\_1” I collapse the sum of “numbirth1544” and group by the data under stname and year. For “state\_births\_2”, I did is everything the same but summarize the “pop1544” instead. Then I merged them in to one dataframe by “stname” and “year”. Then I separate the data into 2 periods and called it year2. The first period is 2004-2008 which I called “2004” and the second one is 2015-2019 and I called it “2019”. After this, I summarised the sum of “numbirth1544” and “pop1544”, groupby “stname”, “year2”. I calculated the rate by numbirth1544/pop1544\*1000 and called it “brate1544\_thsnds”. I generate 2 new data frames for year 2004 and year 2019 and select 3 columns: stname, year2, brate1544\_thsnds for each dataframe. I rename the 3 columns into “state”, “year”, “brate\_bef” for year 2004 and “state”, “year”, “brate\_aft” for year 2019. By function merge I convert two data frames into one final data frame. I added up a new column to record the changes in birth rates between two periods by function mutate and I called it “diff”. By package “usmap” I can do a plot of all 50 states in united states. This figure is called Change in Birth Rates by State between 2004-2008 and 2015-2019. The percent change of birth rate from low to high has different colors. I choose white as low and red as my color of highest. The last step is to zoom out the labels into size 3 to avoid the messy looking on the east coast.

Figure 3 is a state-level US map that indicates the changes in birth rates among women aged 15 to 44 between two five-year periods (2004-2008, 2015-2019) before and after the Great Recession.

Moreover, we calculate the sum of new birth as well as the existing population of each five years period for states and save them in two newly created dataframes. We combined the two datasets by grouping states’ names and years. To simplified, we renamed the years between 2004 and 2008 as “2004” and those between 2015 and 2019 as “2019”. Then, we summarized the sum of new birth and existing population by state and year. The rate of birth among women aged 15 to 44 is calculated by dividing the number of new birth by the number of existing populations multiplied by one thousand. After that, we created two dataframes to separate observations of the year “2004” and year “2019” and kept only state name, year, and corresponding birth rate in the dataframes. To continue, we merged these two dataframes as the final one and recorded the changes in birth rates between two periods (2004-2008, 2015-2019). By combining data packets with maps of the United States, we made a plot of 50 states in the US with percentage change of birth rate and named as “Change in Birth Rates by State between 2004-2008 and 2015-2019”. Specifically, the higher a state’s fertility rate, the redder it is plotted. The color approaches white when the birth rate reaches negative twenty percentage points and approaches bright red when the birth rate is turned to be positive.

## Result

### Trend in US Birth Rates

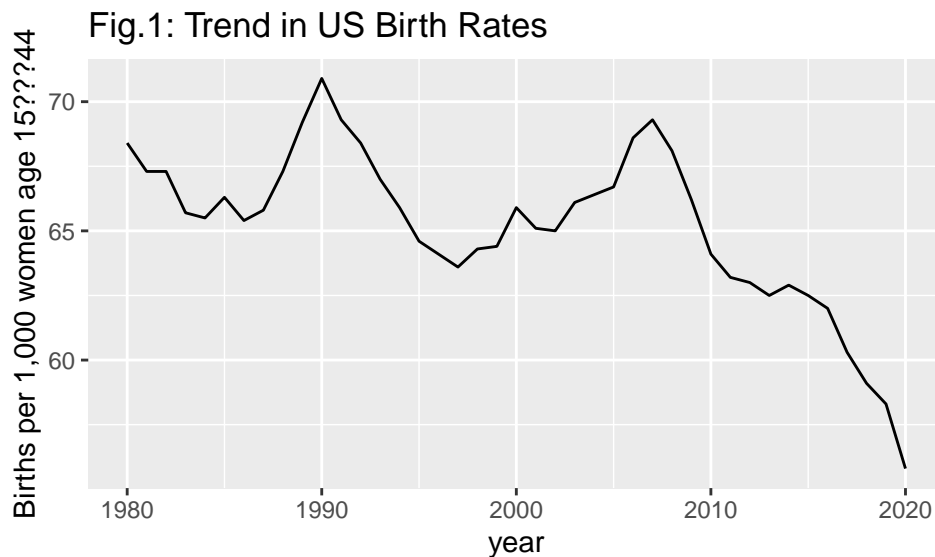


Figure 1 reflects trends in U.S. fertility from 1980 to 2020, using data on the universe of US births from the Vital Statistics system. It can be seen that the overall fertility rate in the United States is on a downward trend, from about 68% in 1980 to about 56% in 2020. During this time, there were two obvious fluctuations, near the year 1990 and 2007, respectively. The decline around 1990 began to pick up around 1996; the fertility decline that began in 2007, however, was sharp and persistent. This paper takes the fertility decline since 2007 as the research subject and aims to explore the potential strong correlation factors behind it. We retrieved major events, including politics, economy, and culture, during this period as potential causes of the great turning point in 2007, and found that the great recession was one of the significant events.

The great recession brought enormous economic stress to people at the time: Between 2007 and 2010, the unemployment rate rose from 4.6% to 9.6%, about 5 percentage points. The appearance of a newborn will bring new expenses to a family, but at the same time, it can also bring economic income as a new labor force for the family in the near future. There is a lot of literature trying to define the economic impact on the birth rate. Dettling and Kearney (2014) mentioned that increased income leads people to choose to have more children. However, high economic prosperity also raises the standards and costs of raising children, such as in education and quality of life. Becker and Lewis (1973) draw conclusions based on this fact that people tend to have fewer children while ensuring high-quality investment in each child, as society grows richer. Therefore, without quantitative analysis, we cannot confirm the specific impact of the Great Recession on U.S. birth rates.

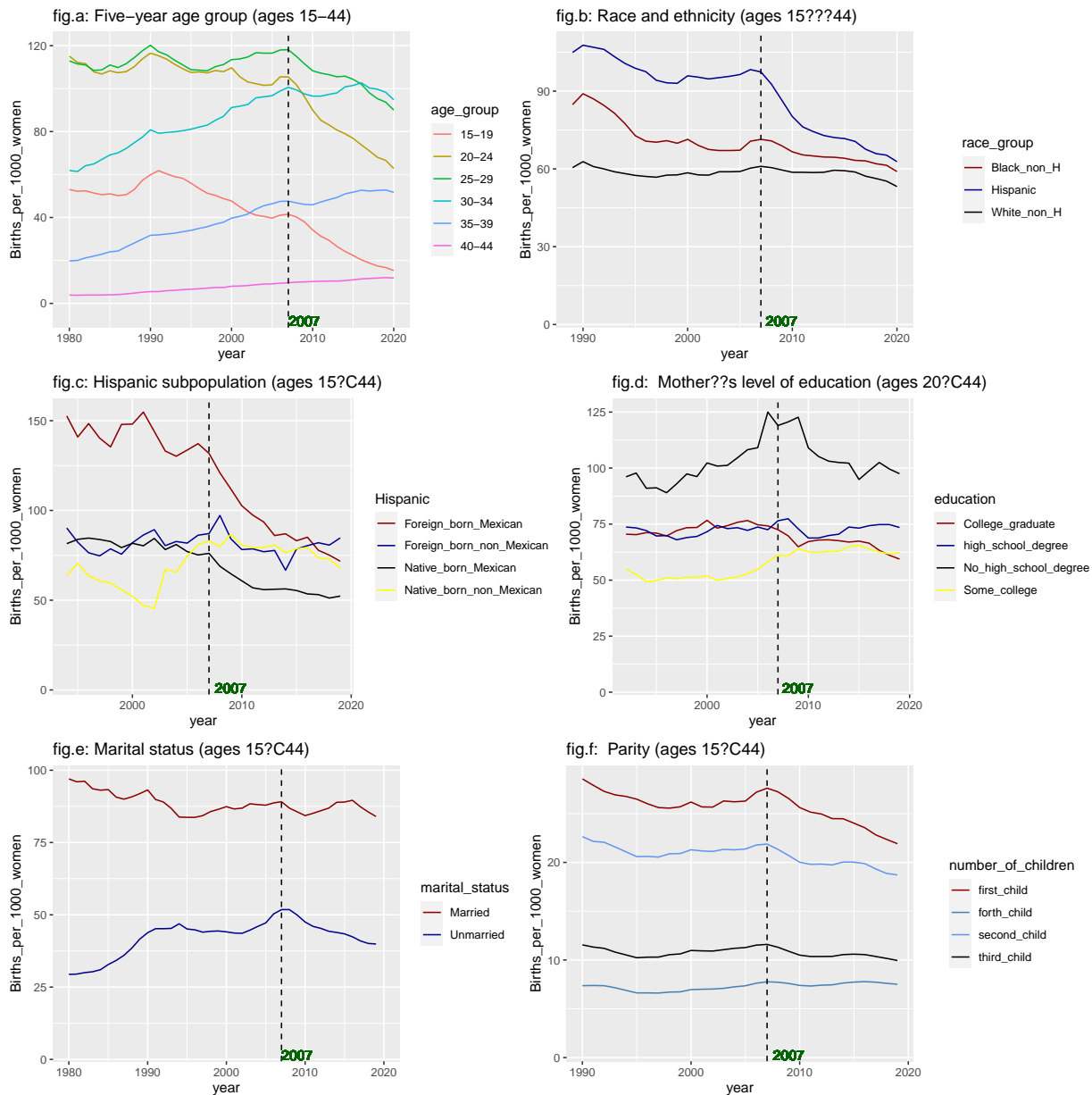
To be sure, however, the birth rate continued to decline after the recession ended. As can be seen from figure 1, after 2010, the birth rate was still falling. Although a rebound occurred in 2013, it is too slight and brief to reverse the overall sharp decline. Therefore, it can be concluded that factors other than the Great Recession played a role in the continued decline in the birth rate since 2007.

### Trends in Birth Rates by Population Subgroup

In Figure 2.A, we draw the change in birth rate of each year from 1980 to 2020, between 6 age groups. The change in births between age group 15-19 is the most distinct pattern in all 6 lines. The births per 1,000 women shows a downward trend after reaching a peak in the year of 1991. At around the year of 2007, the rate of decline slowed for around 3 years before speeding up in the following years. Women from the age groups of 20-24, and 25-29, both experienced a decrease in births. The birth rate of elder women (age 30-34, age 35-39 and age 40-44) rose gradually during the 40 years. However, these increases in birth among elder women are not even close to compensate for the significant reductions in births in younger age groups. With the increasing average levels of education become higher, society is more prosperous, young people know more about contraception, and the rate of adolescent pregnancy has fallen (10.1111/jomf.12815)

In panel B of Figure 2 investigates birth rates by race. We see from the graph that the birth rate of Hispanics experienced the most significant drop. The birth rate of Hispanics drops from 97.4 in 2007, to 62.8 by 2020. Birth rates for non-Hispanic Black and White women declined as well, but by considerably lower percentages. When the Great Recession hit, race and ethnicity had a significant impact on birth rates. By 2020, racial and ethnic inequalities in birth rates will still exist, but they will be significantly reduced. The graph shows a pattern of fertility convergency among ethnicity after 2007. Fertility convergence is strongly associated to educational convergence. The significant improvements in educational attainment across immigrant generations have become a driving factor of fertility convergence among ethnicities (Intergenerational Fertility Among Hispanic Women). A possible explanation to this phenomenon is that the Great Recession resulted in educational convergence among races, which further led to fertility convergence among Hispanic, non-Hispanic Black and White.

From Figure 2 panel C, it's not hard to see that Hispanic was the group that fertility rate is decreasing the most. Thus, we analyzed the trend of the Hispanic subgroup separately in panel C. Specifically, we examine birth rates among native-born and foreign-born Hispanics separately, as well as Hispanics by country of origin, Mexican and non-Mexican. The birth rate of foreign-born Mexican and native-born Mexican both drop a large extent. While the birth rate non-Mexican Hispanic women only fluctuate in a small range. The hard drop in Mexican contribute to the significant decline of Hispanic subpopulation birth rate. The figure shows that the birth rate of foreign-born Mexican is higher than that of native born Mexican, which supports the assimilation.



Following that, we'll look at birth rates for four different maternal education groups: those with less than a high school diploma, those with a high school diploma, those with some college diploma, and those with a four-year college diploma or more. Because many teenagers have not completed their education, we excluded them from this study. The most recently educated and least educated women, women with four-year college degrees (36.3 percent of women aged 20 to 44 in 2018) and women without a high school diploma (8.1 percent of these women), saw the highest decreases in birth rates, as seen in Group D in Figure 2. Between 2007 and 2019, the birth rate for college-educated women declined from 72.5 to 59.4 per 1,000, while it fell from 119 to 97.5 for women without a high school diploma. Since 2007, both groups of women, with a high school diploma and those who have gone to some college, have had very constant birth rates.

We then want to investigate if marital status of mothers would influence the birth rate. Panel E of Figure 2 separately plots births by two maternal groups, married and unmarried. From 1980 to around 1994, the births of two maternal groups experienced substantially different trends. The births of both groups of women

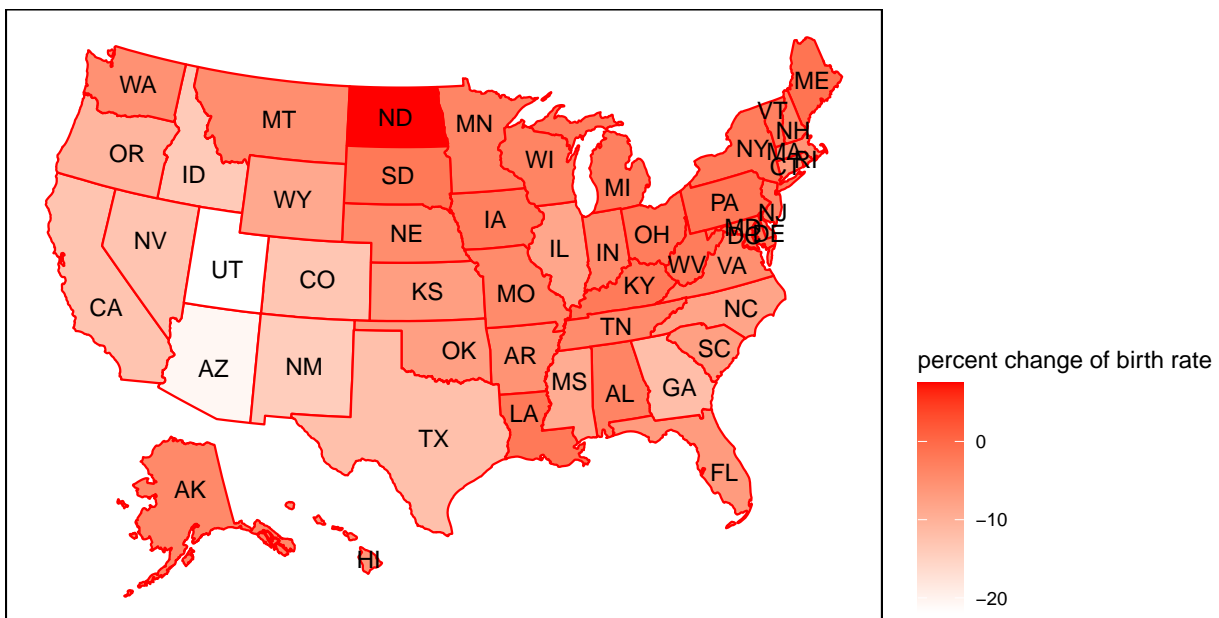
slightly flatulated between the year of 1995 to 2007. Nevertheless, between 2007 and 2010, the birth rate of married women and unmarried women went down by similar extents. Starts from 2010, the births given by married women in the US society rose again after a fall. However, the births of unmarried women have been dropped continuously since 2010. After 2007, although there exists difference between the change of birth rate in the between years, the overall patterns of the two maternal marital status groups are similar, each experiences a small drop. Hence, it is eligible to conclude that marriage status only influences birth rate in a small range.

Panel F of figure 2 depicts parity of the number of women who give birth in per 1,000 women, where parity refers to the number of times that a specific woman has given birth to a fetus. The line of the First birth has the most significant change, while the line for women who have their fourth or more birth, which suggests that the drop in births after 2007 is due to a decrease in starting to have kids rather than women not willing to have larger families.

To further explore if the above attributes indeed contribute to the drop of birth rate, we try to show the trend of change in birth rate by geographical division.

### Change in Birth Rate by State, 2004-2008 to 2015-2019

Fig.3: Change in Birth Rates by State between 2004–2008 and 2015–2019



There are many reasons for the decline in fertility, such as the increase in the effectiveness of contraceptive methods, the increase in social competition pressure, and the improvement of women's social rights. But which of these could explain the long-term and persistent fall in birth rates of the US after 2007 remains to be determined. Figure 3 depicts the state-by-state trends in the U.S. population, measuring changes in birth rates by calculating population changes over two five-year periods before and after the great recession (2004-2008, 2015-2019). By taking five-year averages to measure population status and calculating birth trends across more than 10 years, we can somewhat circumvent reasons for short-term fluctuations.

As shown in the Figure 3, the overall fertility rate in the United States is on a downward trend. Except for the positive fertility growth rate in North Dakota, all other states are in negative growth. Fertility declines in most states ranged from zero to five percentage points, followed by declines of more than ten percentage points. Among them, the overall decline in fertility in the western and southwestern states of the United States was larger, with OR, ID, CA, NV, UT, CO, AZ, NM, and TX fertility declining by more than ten percentage points. By contrast, the northern and eastern states saw smaller declines in fertility, mostly zero to five percentage points decline. And one possible explanation for the rising birth rate in North Dakota is that the state has experienced a fracking boom over the years, which other studies suggest has increased birth rates (Kearney and Wilson 2018).

The Southwest and Western states saw the biggest drop in birth rates. In much of the region, the Hispanic population is sizable, consistent with a sharp drop in the birth rate for Hispanic women. We know from Figure 2 that this is due to a decline in the birth rate of native and foreign-born Mexicans.

Thus, the importance of declining birthrates for Hispanic women is corroborated by state-year analysis as a potential long-term and slow meaningful effect, rather than a short-term disturbance, as one of the reasons for the continued decline in U.S. fertility rates after 2007. It is worth noting that, according to the conclusions drawn from Figure 2, Hispanic women's fertility has experienced convergence, and the study of EMILIO A. PARRADO AND S. PHILIP MORGAN shows that education convergence is the main reason behind the fertility convergence. Therefore, we can infer that education change can be considered as a potential reason for the decline in the US fertility rate after 2007.

## Discussion

### Discussion of Research Findings

#### Weakness

1. Lack of discussion of population base for Hispanic women

We have examined the magnitude of changes in fertility rates for Hispanic women. Educational convergence is a key reason behind the dramatic decline in birth rates among Hispanic women due to assimilation toward native non-Hispanic American women. In addition, through state-to-state analysis, we found that fertility changes among Hispanic women have long-term and regional effects on U.S. birth rates after 2007. Therefore, it is clear that the fertility rate of Hispanic women and the educational convergence behind it may be one of the reasons for the continued decline in the birth rate in the United States after 2007.

However, the impact of a subgroup on the overall population depends not only on its magnitude of change, but also on its population base. We did not examine the percentage of Hispanic women of childbearing age in the United States after 2007. In fact, despite the plummeting birth rate among Hispanic women, they make up only 16.7 percent of the population. So, observed from a mathematical perspective, their impact on overall fertility would be greatly reduced. While we acknowledge the persistence and impact of this factor, it is uncertain whether the decline in the birth rate among Hispanic women and the educational convergence behind it will lead to a sharp decline in the birth rate in the United States after 2007. In other words, it is only eligible to say that it will drop the US fertility rate after 2007, but the extent of the effect is undetermined.

- 2.

In Fig.3 the todo list in readme file has different values as it shows in the final product report. The readme indicates the second period as 2015-2018 but the report shows the time period between 2015-2019. If we use 2015-2018 instead of 2015-2019, the difference between the rate of change of first period and second period could be larger.



## **Future Work**

To determine the contribution of the decline in the fertility level of Hispanic women to the continued fall in the overall fertility rate in the United States after 2007, we can further establish statistical models, such as the linear regression model, for quantitative analysis. As a replication study, this article can assess the modeling method used in the original paper, that is, to classify race, age, and education level, and design a total of 63 demographic groups. On this basis, we can focus on the Hispanic fertility rate, study its contribution proportion to the whole, as well as its characteristics in various age stages and educational backgrounds.