

A Dramatic Drop of US Birth Rates after the Great Regression*

Factors are found, but not enough evidence to explain the reason behind it

Zhiyue Yu

Dingding Wang

Xueru Ma

28 February 2022

Abstract

Birth rate is one of the most important indicators that monitors community health. When the birth rate falls below replacement level, the population grows older and shrinks, a large gap then occurs in the labor market, thus slowing the economic growth and strains the government budget(Murray, 2021). The data we used for research and analyzing is collected through a research paper published by the American Economic Association; the data sets are published on OPENICPSR(Kearney and Levine, 2022). After plotting a few graphs and applying a linear regression model, We find that there is a strong association between the decline in birth rate after 2007 and there are a few elements that dominate the decline in birth rate, but we cannot find strong evidence to explain why these factors cause this change. Our findings provide valuable suggestions for the government and organizations to deal with aging population and shrinking workforce challenges.

Contents

1	Introduction	1
2	Data	2
2.1	Overview	2
2.2	Demographic	2
3	Model	4
3.1	Linear Regression	4
3.2	Full model, Reduced model and ANOVA	4
3.3	Conditions and Assumptions Check	6
4	Results	8
5	Discussion	8
5.1	Changes of US Birth rate before and after the Great Recession	8
5.2	The Overall Change in the US Birth Rate	9
5.3	Limitations	11
	References	12

1 Introduction

Birth rate is one of the most important indicators that monitors community health. When the birth rate falls below replacement level, the population grows older and shrinks, a large gap then occurs in the labor market, thus slowing the economic growth and strains the government budget(Murray, 2021). According to the New

*Code and data are available at: <https://github.com/gggeorgeeyu/sta304-paper2>. The DOI is: 10.1257 / jep.36.1.151 see the article page at <https://doi.org/10.1257/jep.36.1.151>.

York Times, the COVID-19 pandemic is accelerating the decline of US birth rate after 2020 (Tavernise, 2021), however, it is not the reason for the birth rate decline after the Great Recession in 2007. In the period of 1980 and 2007, the birth rate of the United States fluctuated between 65 and 70, whilst there was a dramatic drop after 2007(Figure 1), so what is the reason and explanation after this phenomenon?

The data we used for research and analyzing is collected through a research paper published by the American Economic Association; the data sets are published on OPENICPSR(Kearney and Levine, 2022). The data shows all the demographics and geographics of the US population, as well as other miscellaneous elements related to the population, such as education, economics, etc..

After plotting a few graphs to explore the demographic and applying a linear regression model to it, We find that there is a strong association between the decline in birth rate after 2007 and there are a few elements that dominate the decline in birth rate, such as birth rate of mother aged from 20 to 24 and aged from 30 to 34, birth rate of married mothers, etc., but we cannot find strong evidence to explain why these factors cause this change. Our findings will provide valuable suggestions for the government and organizations to deal with aging population and shrinking workforce challenges. We provided some of our opinions in the Discussion(Section ??).

2 Data

In the paper, we run our analysis in R (R Core Team 2019). To plot all the line charts we needed to analyze, we used `tidyverse` package written by (Wickham et al. 2019), `ggplot2` written by (Wickham 2016). When processing the raw data sets, we used `haven` package written by (Wickham and Miller 2020). To open dta files we used `dplyr` package written by (Wickham et al. 2021) to apply filter function, to open csv files, we used `readr` package written by (Wickham and Hester 2021). To arrange the plots, we used `patchwork` writted by (Pedersen 2020) We also used `usmap` written by (Di Lorenzo 2022) to draw the us map.

2.1 Overview

Figure @ref(fig: fig1) shows the tendency of the birth rate in the United States. It is observed that the decline in the birth rate began from the Great Recession and has continued to decline even after the economy recovered over time. We can view this graph in two parts, one prior to 2007 and a part after 2007. Before 2007, we found that the birth rate changes as if in a cyclical pattern, and that the birth rate fluctuates between a narrow range of 65%-70%.The birth rate has a general downward trend between 1980 and 1987. Subsequently, the birth rate gradually rose and it reached a peak in 1990. Similarly, in the interval from 1990 to 2007, the birth rate first decreases and then gradually increases, and it reaches a relatively high level in 2007. The birth rate at this time was close to the previous maximum birth rate in 1990. However, the state of change in the birth rate after 2007 gradually differs from our previous speculations. We find that the birth rate continues to decline over the thirteen years from 2007 to 2020, and that this decline spans a much longer period of time than the length of the part of the cycle we previously estimated to be declining (five to seven years). At the same time, we also notice that the magnitude of the decline is very large. The birth rate plunged from 69 percent in 2007 to 56 percent in 2020. The change in the birth rate no longer lies within the 65%-70% range, but falls below it at a rapid pace.

2.2 Demographic

2.2.1 figure 2a

To figure out what are the factors that influence the birth rate most, we explored several possible factors. In figure @ref(fig: fig2), we can tell that people in the age group of 25-29 have the highest birth rate, followed by that of people in the age group of 20-24. Apart from the line charts of people in age group 30-34 and 40-44, which rose after 2007, the rest of the age groups have a clearly declining birth rate after 2007. Among those groups, people in the age group 20-24 and 25-29 may have the most influence in the overall birth rate, because they have the top two highest birth rates.

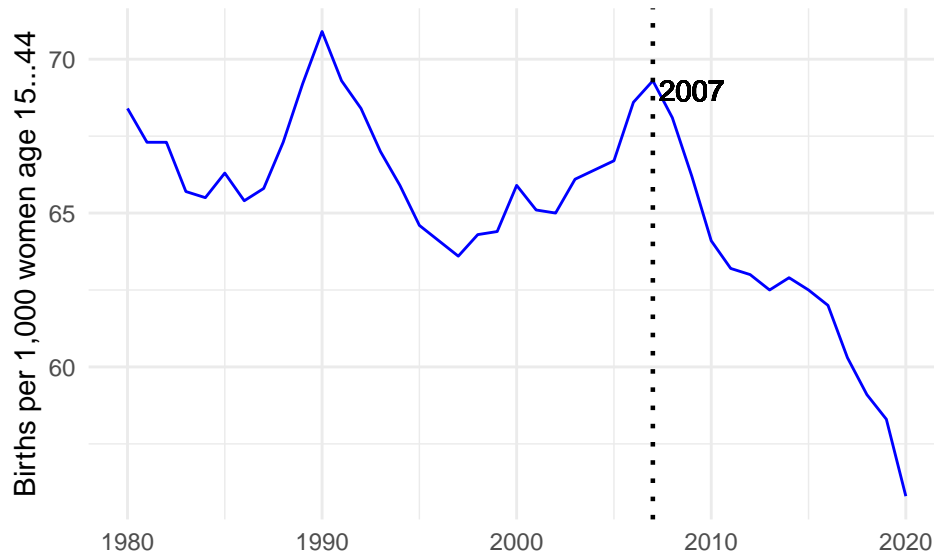


Figure 1: Trend in US Birth Rates from 1980 to 2020

2.2.2 figure 2b

In figure @ref(fig: fig2), we explored the effects of race and ethnicity. From the graph, we can easily tell that hispanics experienced a dramatic drop after 2007 from 97 to 62. For those non-hispanics, both black and white people have dropped after 2007, but it is a slight drop and doesn't make a big difference. In this graph, we missed minority groups such as asian(Chinese, Korean and Japanese, which also occupies a large group of people and this is a limitation for the data set and the graph.

2.2.3 figure 2c

In Figure @ref(fig: fig2), we can easily see that only the foriegn born Mexican have had a dramatic drop after 2007, from 130 to 70, which is a great contributor to the overall drop of US birth rate. However, for the other groups of people such as Native born Mexicans, there is only a slight drop from 75 to 50. The rest of the hispanic population have zero or negative impact on the drop of birth rate of US hispanic population.

```
## Rows: 51 Columns: 23
## -- Column specification -----
## Delimiter: ","
## dbf (23): year, tfr, brate_all, brate_1519, brate_2024, brate_2529, brate_30...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

2.2.4 figure 2d

In figure @ref(fig: fig2), we can see that people with no high school degree tend to have a much higher birth rate and a much more fierce drop of birth rate after 2007, this may due to the impact of Great Recession on their incomes and their mind changes of having a kid. For people with high school and higher degrees, they are affected less by the Great Regression. One possible reason is that they have more deposits for emergencies when the whole world is under financial crisis, so they will not have a big change with their initial plan when it comes to having a baby.

2.2.5 figure 2e

In figure @ref(fig: fig2), we explored the factor of marriage. We can tell that married people have a birth rate of around 95, which is much higher than that of unmarried people, whose birth rate is only about 55. Both of the groups did not have a dramatic drop after 2007. This aligns with our common sense and actually did not provide further insights.

2.2.6 figure 2f

In figure @fig(fig: fig2), we tabulated the trend of birth by parity. It is obvious in the graph that the first birth is affected most by the Great Recession after 2007 where there is a dramatic drop from a rate of 27.6 per 1,000 women of childbearing age to 21.9 per 1,000 women. The effect declines when the number of births increases, the fourth plus birth is rarely affected by the Great Recession. This may be due to the reason that people with less education tend to have more children and their income are not affected by the Great Recession that much, thus will not have an impact on the birth of the number of children.

3 Model

3.1 Linear Regression

Regression analysis, both linear and multivariate, is a widely-used statistical method used to quantify the relationship between one dependent variable and the other independent variables that are thought to explain it. Regressions can also identify how close and well determined the relationship is. A Multivariate regression model has been used on various indicator variables using a database collected from both public sources and aggregated from the National Center for Health Statistics' (NCHS) microdata that are restricted in use(Kearney, Levine, and Pardue 2022). In this case, the independent variables X include the important variables such as age groups, hispanic origin, marital status, and state. The final interpretation of the rate of birth can be estimated from the observed values of the independent variables with appropriate reduced models.

The method for our model selection is forward selection method, a type of stepwise regression that uses statistical significance to select the explanatory variables, which starts with an empty model and adds variables one by one. The goal of stepwise regression is to find a set of independent variables that significantly influence the dependent variable(Birth rate) through a series of tests, such as F-tests and t-tests(Hayes 2022). This is done with iteration, as We add the predictor variables one by one and examine the SS_{res} , R^2 , R^2_{adj} and AIC at each stage of the model. The determination of whether the addition of variables improves or worsen our model is by conducting examinations of whether the values of SS_{res} , AIC, BIC have decreased, along with whether the R^2 and R^2_{adj} have increased. In the adjusted model, 2 non-significant predictors were removed from the full model, remaining only 2 age groups, race, marital status and Hispanic origin.

3.2 Full model, Reduced model and ANOVA

By applying ANOVA test, the result shows a non-significant result ($p = 0.5399$). Thus, we should reject full model and stick with reduced model

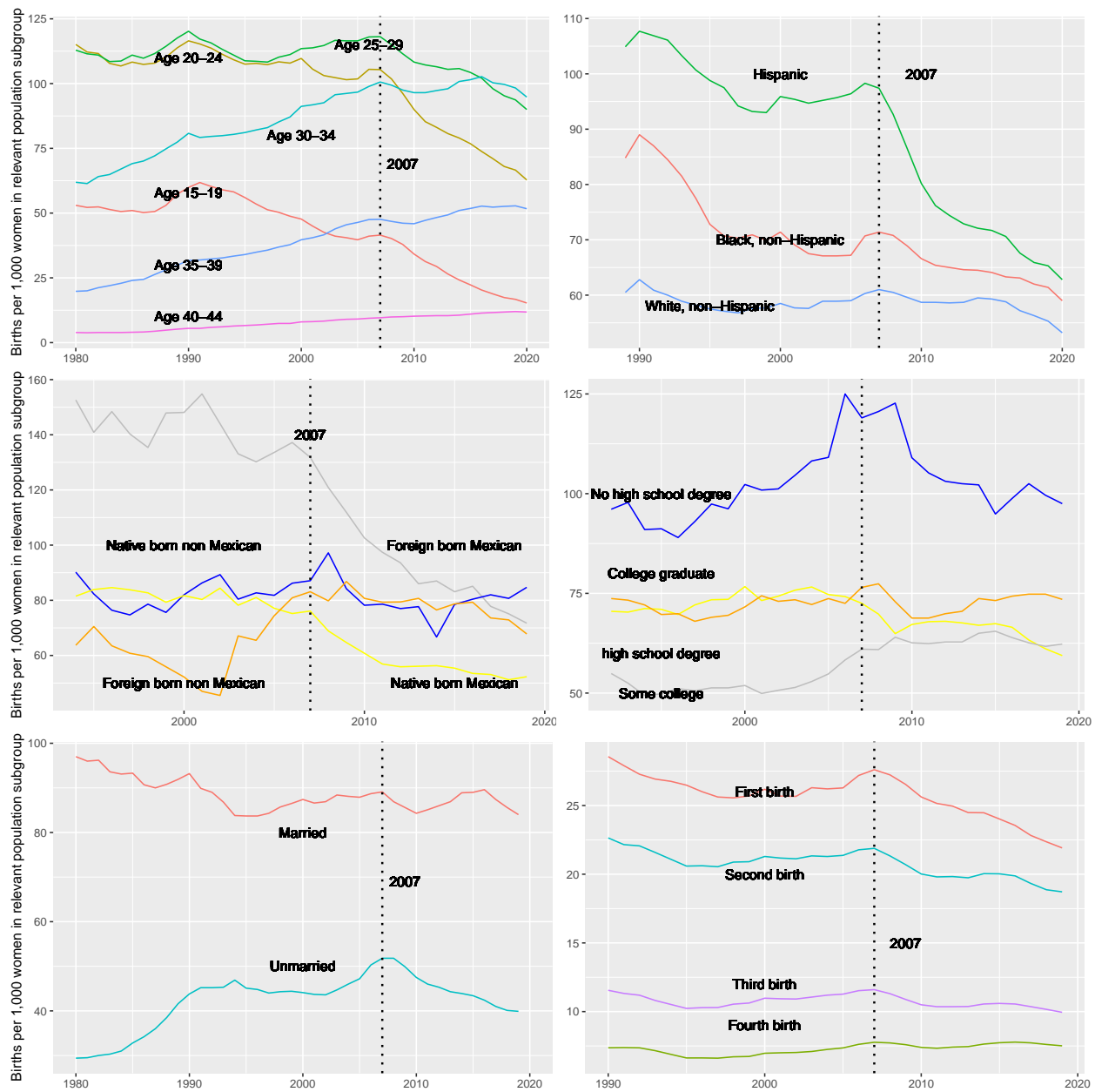
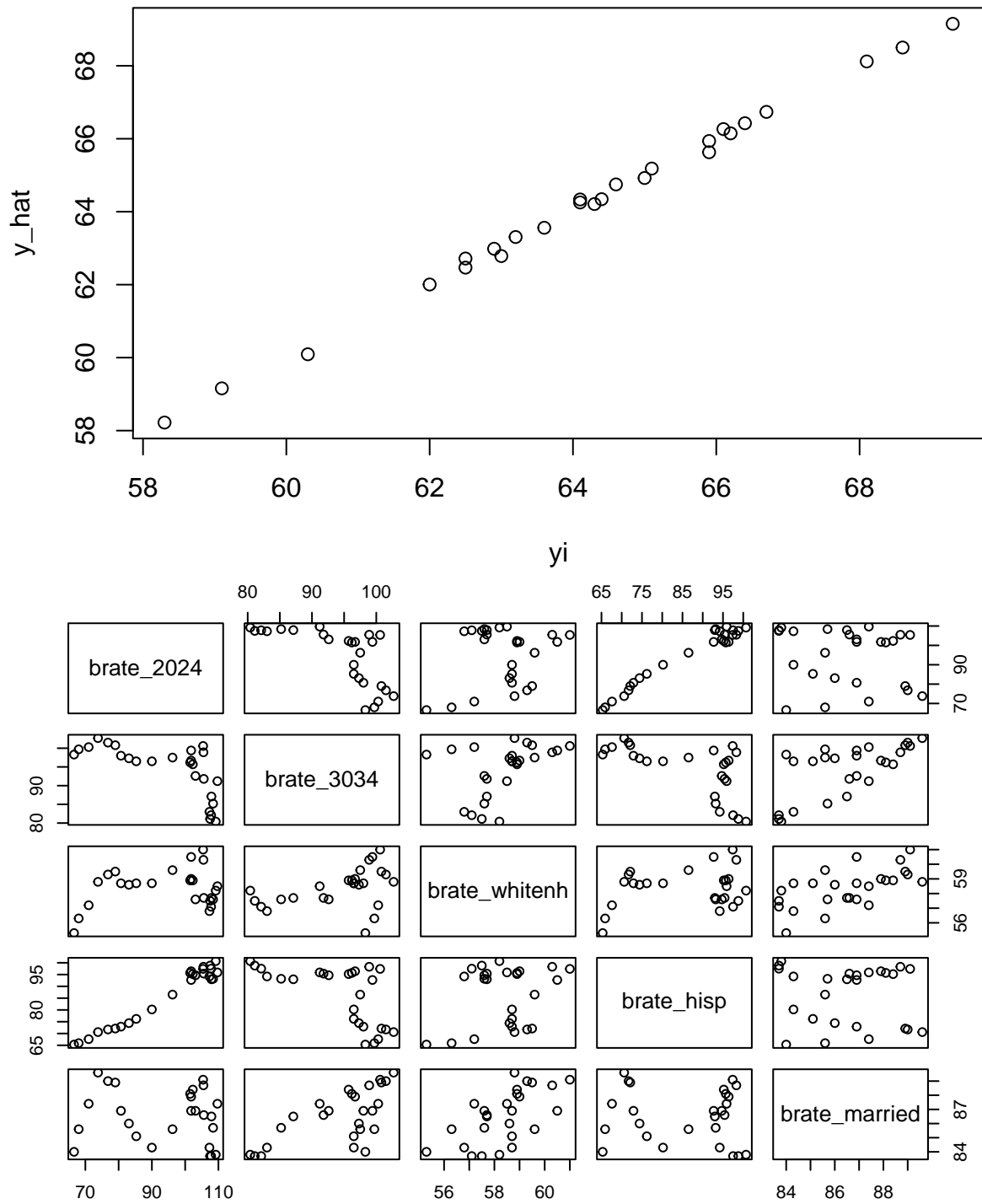
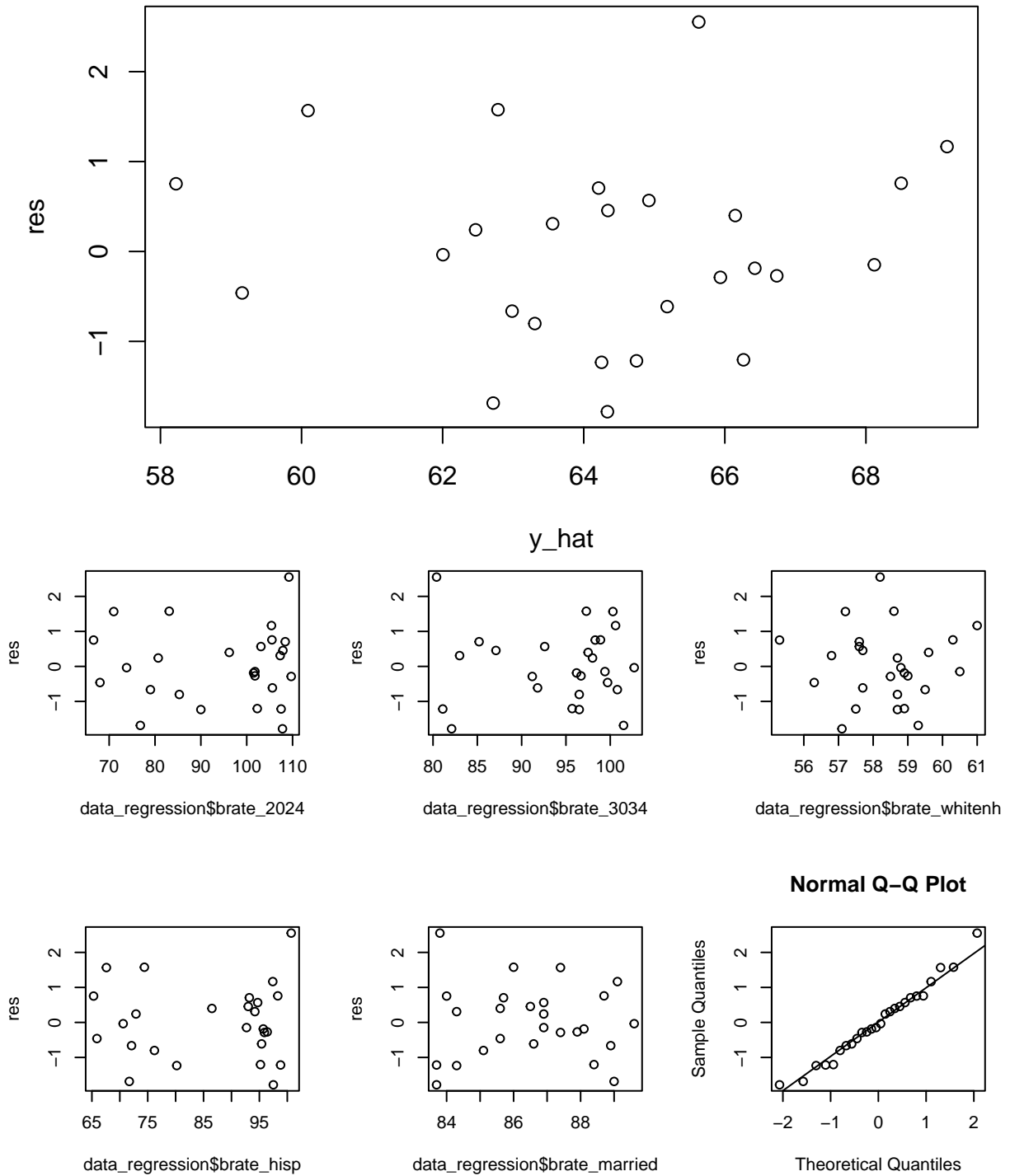


Figure 2: Trends in Birth Rates by Population Subgroup

3.3 Conditions and Assumptions Check





For a linear regression model to make sense, it should satisfy the below conditions:

- Relationships between each pair of quantitative X and Y are straight enough.
- The residuals across fitted values are about equal spread.
- No extreme outliers exist. We want to avoid extreme outliers since the points in the model that do not fit the overall relationship and are far away from the mean value of X have leverage to change the line, leading to a biased result.

We have to examine how well the above conditions are satisfied when we build our regression model. To

check the first two conditions, we take a look at both the original scatterplot and the plot of the residuals against fitted or predicted values. From the plots,(Figure @ref(fig: linear)) no sign of violation of normality or constant variance assumptions is observed, and the normal QQ plot suggests that the residuals follow a right-tailed distribution. The regression errors also have a normal distribution.

Adjusted R-squared	Degrees of Freedom	p-value
0.40247	2	0.5399

Our final model interpreted is: Total Birth Rate = $-1.40503 + (0.04456)brate_2024 + (0.13034)brate_3034 + (0.91645)brate_whitenh + (0.14543)brate_hisp + (-0.19442)*brate_married$

4 Results

In this paper, we replicated the paper published by the American Economic Association by plotting line charts to dig the contributing factors to the drop of birth rate and drawing a usmap to demonstrate the distribution of the change of birth rate before and after the Great Recession. We also applied a linear regression model to the processed data sets to find the predictors with systematic statistical methods and we successfully figured out a reduced linear regression model that explains the drop in birth rate. Although we have applied what we learned to explore the reason behind the “US birth rate dropping” phenomenon, we can only find the factors that affect the result, however, we do not have enough evidence to explain the reason behind those contributing factors. The reason behind the simple social phenomenon may be complicated: it may be due to the economics, social norms, the distribution of resources, or maybe the mindset of having a kid has changed, young people tend to pursue a lifestyle with more freedom, thus they prefer a life without having kids. With whatever reason for the birth rate drop, we have to face the result of the birth rate declining – not only for the US, which leads to an aging population and shrinking workforce, posing challenges for economic growth and sustainable development of the society. To address this issue, the government may establish a better welfare system to reduce the cost and pressure of having a child, specifically free education, free health care and enough mental health support. The government should also cover more daily expenses of new born families and build more day care institutions to reduce the pressure of parents from work and life.

5 Discussion

5.1 Changes of US Birth rate before and after the Great Recession

Figure @ref(fig: fig3) is a map that illustrates the change in birth rates for each state in the United States between 2004-2008 and 2015-2019. We know that 2004-2008 is the time period prior to the Great Recession, while 2015-2019 is the period after the Great Recession. By looking at the change in birth rates for each state over the same five-year period, we can capture the impact on each state’s birth rate as a result of the different economic and political policies adopted by each local government during and after the Great Recession. As we all know, each state has a different economic priority, with some states having a predominantly agricultural economy and others having a more business-oriented economy, so the Great Recession had varying effects on them. After a recession, governments often introduce different policies in order to recover the economy more rapidly. What we see here is actually the effect of the negative effects of the Great Depression offset by the positive effects of the recovering policies. Here we do not discuss the birth rate for the entire country, but instead break it down to each state, in such a way that avoids the potential bias when considering different situations everywhere with the averages. Such an approach provides a more accurate reflection of each state’s unique situation. The entire map is divided into many regions on a statewide basis, and the colors of the regions represent the change in birth rates in those states. The darker the color of the filled area is, the more severe we find the reduction in birth rate in that area, and the lighter the color of the filled area is, it indicates that the Great Recession has a less significant impact on the birth rate in that state. In general, almost every state experienced a decrease in birth rates after the Great Recession. Thus we can get the conclusion that the Great Recession has a negative impact on the birth rate. However, at the same time,

we also find that birth rates in North Dakota actually increased after the Great Recession. We think it is possible that this is the migration Response to the Fracking Boom(Wilson 2020). Exploiting positive labor market shocks from localized “fracking” booms, the birth rate in North Dakota is supposed to increase. Meanwhile, we observe that the two states, Utah and Arizona, had the most serious impact on birth rates from the Great Recession.

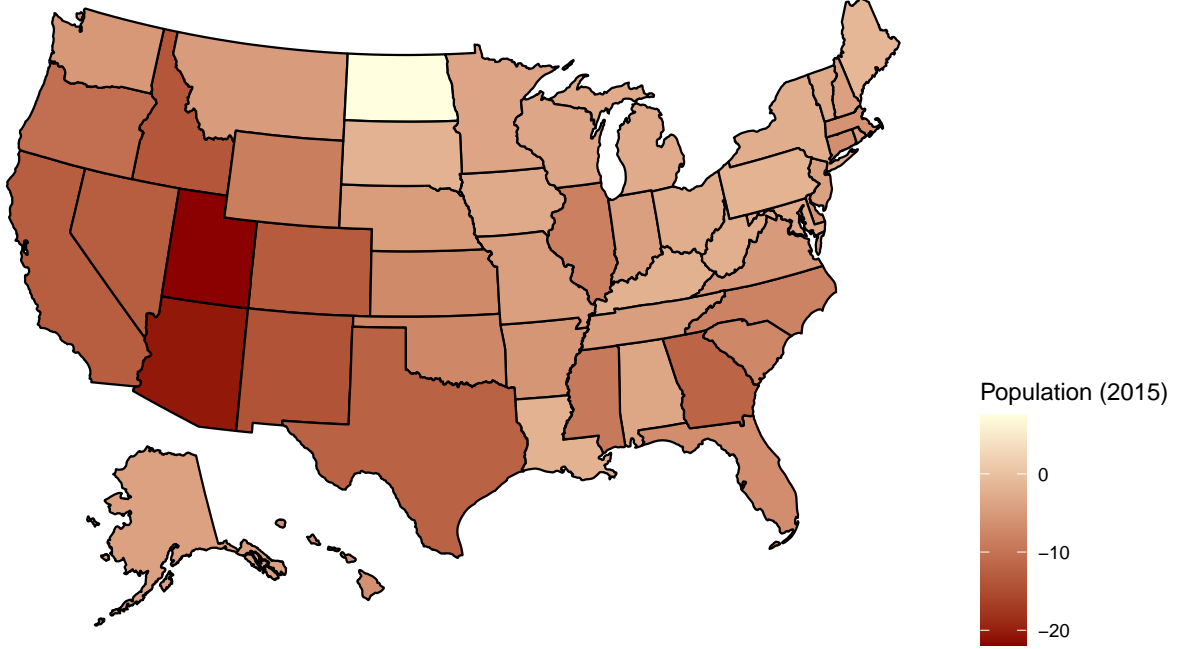


Figure 3: Change of Birth Rates by State from 2004–2008 to 2015–2019

5.2 The Overall Change in the US Birth Rate

The overall change could be written as the combination of the contribution of changing within-group birth rates, the contribution of changes in group population shares, and the interaction of a group’s changing rates and changing population shares.

$$\Delta \left(\frac{B}{P} \right)_{t_0, t_1} = \sum_i s_{i, t_0} \Delta \left(\frac{B}{P} \right)_{i, t_0 t_1} + \sum_i \left(\frac{B}{P} \right)_{i, t_0} \Delta s_{i, t_0 t_1} + \sum_i \Delta s_{i, t_0 t_1} \Delta \left(\frac{B}{P} \right)_{i, t_0 t_1}$$

where: $-B$ represents the amount of births

$-P$ represents the population

$-s$ represents the overall population

$-i$ represents the index among the 63 groups

$-t_0$ represents the beginning year

$-t_1$ represents the ending year

To transform this to a table:

variable name	meaning
B	the amount of births
P	the population

variable name	meaning
s	the overall population
i	the index among the 63 groups
t_0	the beginning year
t_1	the ending year

5.3 Limitations

5.3.1 Stepwise Regression

As a combination of the forward and backward selection techniques, stepwise regression is widely used in different areas of regression analysis. Stepwise regression requires two significance levels, one is for adding and another one is for removing variables from models. However, a fundamental problem with stepwise regression is that some truly explanatory variables that have a causal effect on dependent variables may happen to be statistically insignificant, while annoying variables may happen to be significant. Therefore, the model fits the data well in sample, but poorly out of sample (Smith 2018). In practice, the stepwise regression process may sometimes select explanatory variables that, although they do not directly affect the dependent variable, are systematically related to variables that do. In our study, for example, fertility depends on marriage rates, which in turn correlates with income. Even if income does not directly affect fertility, it is a proxy for income and may find its way into the stepwise regression equation. In reality, stepwise regression fails to solve the problem of Big Data with too many explanatory variables, making it less effective compared with other model selection techniques.

5.3.2 Limitation of Data

It is worth noticing that the data used in this publication are estimates based on data collected and are subject to error. There are two types of errors: sampling errors and non-sampling errors. Sampling error occurs due to the sample selected that fails to represent the population, and a non-sampling error refers to an error that results during data collection process, which causes the data to differ from the true values (Kenton 2022). In the study above, we describe some survey evidence related to our hypothesis that shifting priorities across cohorts may be a key explanation for the declining birth rate in the United States. We need to be aware of the limitations of self-reported survey responses. Self-reports on priorities, attitudes, reasons, etc., are often subject to interpretation, sensitive to survey wording or current context, plagued by recall issues, and 31 other limitations. In addition, the paper uses a limited amount of survey evidence, especially those that have been repeatedly asked to track trends for a long time.

5.3.3 Population Base for Hispanic Women and the effect

In our linear regression model, the fertility rate of Hispanic women is an important variable. However, when considering the effect of a particular race's fertility rate on a country's overall fertility rate, we should also consider the racial percentage of the country's total population. Latinos accounted for 18% of the US population, or almost 59 million people as of 2017. Thus, in proportion to the population, the extent to which Hispanic women have contributed to the precipitous decline in the birth rate in the United States is not obvious. However, it is not yet certain that the declining birth rate among Hispanic women and its potential impact will lead to a sharp decline in the birth rate in the United States.

References

- Di Lorenzo, Paolo. 2022. *Usmap: US Maps Including Alaska and Hawaii*. <https://CRAN.R-project.org/package=usmap>.
- Hayes, Adam. 2022. “Stepwise Regression.” *Investopedia*. Investopedia. <https://www.investopedia.com/terms/s/stepwise-regression.asp>.
- Kearney, Melissa S., Phillip B. Levine, and Luke Pardue. 2022. “The Puzzle of Falling US Birth Rates Since the Great Recession.” *Journal of Economic Perspectives* 36 (1): 151–76. <https://doi.org/10.1257/jep.36.1.151>.
- Kenton, Will. 2022. “Non-Sampling Error.” *Investopedia*. Investopedia. <https://www.investopedia.com/terms/n/non-samplingerror.asp>.
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Smith, Gary. 2018. “Step Away from Stepwise - Journal of Big Data.” *SpringerOpen*. Springer International Publishing. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0143-6>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain Fran?ois, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain Fran?ois, Lionel Henry, and Kirill M?ller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, and Jim Hester. 2021. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Wickham, Hadley, and Evan Miller. 2020. *Haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files*. <https://CRAN.R-project.org/package=haven>.
- Wilson, Riley. 2020. “Moving to Economic Opportunity: The Migration Response to the Fracking Boom.” *Journal of Human Resources*, 0817–8989R2.