Mathematical Interpretation of China's Economic Outputs

The Role of Birth Rate in Economic Growth

Jianan Liao, Qihang Xu, Ye Liu, Yichen Qian

NEW YORK UNIVERSITY

Supervised by Prof. Timothy Christensen

04/26/2022

Abstract

In this research, we looked for a relationship between China's birth rate and its aggregate production. Based on preliminary findings, we examined the relationship with a focus on evaluating the effectiveness of China's Family Planning Policy in maintaining economic growth. Our initial hypothesis is that China's large labor force supplied by its population is crucial in driving the nation's economics growth. We used panel data regression and fixed effect model to analyze the relationship between China's birth rate and its aggregate production, in hope for the results of this study to provide valuable insights into the role of birth rate in maintaining China's current economic growth and whether Family Planning Policy or other population policies are effective in preserving the basis of a nation's labor force.

Acknowledgement

This report and the research behind it would not have been possible without the exceptional support from our supervisor, Professor Timothy Christensen. His enthusiasm, knowledge, and exacting attention to detail have been an inspiration and kept our work on track from the initial sketch of our research to the final draft of this report.

We also appreciate the funding from College of Arts and Science, New York University. We received the Dean's Undergraduate Research Fund in Fall 2021. The support from the College of Art and Science motivates us to utilize our knowledge from the courses in real-world applications.

Contents

1 Introduction **4**

2 General Assumption **5**

2.1 Assumption I 5

2.2 Assumption II 5

3 Data **5**

3.1 Data Collection 5

3.2 Test for homocedasticity and autocorrelation 6

4 Methodology **7**

4.1 Pool Straight-up Regression 8

4.2 Fixed-Effects Model 9

4.3 Panel Data Regression 9

4.4 Statistical Analysis 10

5 Discussions **10**

5.1 Findings 10

5.2 Limitations 13

5.3 Conclusion 14

5.4 Recommendation 16

1 Introduction

Since the One-Child policy was put into effect in 1979, the pace of population growth in China has slowed down. Despite the effort of maintaining the population growth rate by establishing the "Universal Two-Child Policy" in 2016, the population still reached a minimum value of around 480 thousand in 2021. As a developing country, China's boisterous economic improvement cannot be separated from its population, that China's success in generating huge gross output is partly due to the large workforce derived from its population. However, it is projected that a significant portion of the Chinese population will age to their retirement, resulting in a smaller workforce. Past articles have identified that the proportion of the population over 65 years of age will exceed 18 percent by 2030[ZH16]. The declining birth rate and the aging Chinese population have created tremendous downward pressure on the consistent economic growth and industrial production of China, which concerns the authority. In fact, such concern isn't limited to China alone, other countries that are experiencing population aging and shrinking workforce have established policies such as delaying the retirement age, providing support for families with newborns, etc.

Various policies are proposed by the CCP to alleviate the pressure. In seeking to mitigate the concerning effect that population has on its economy, China has increased the legal age for retirement, assured more rights for women in workspace to increase female participation in the labor force, etc. This report focuses on the effect of China's Family Planning Policy, the most recent being the Political Bureau of the Central Committee of the Communist Party of China proposing the three-child policy on May 31st, 2021. The primary purpose of this report is to conduct a mathematical interpretation sophisticatedly of the impact of China's labor force size on its economic growth. We aim to come to a conclusion regarding the impact of labor force size on China's economy and then analyze the effect of current policies on moderating economic pressure.

2 General Assumption

2.1 Assumption I

In this study, the first assumption we made is that the size of the population plays a crucial role in positively impacting production growth because the human capital it provides is a critical variable in any production. This assumption is based on Adam Smith's arguments, in which he states regards "growth of population as at once a consequence and a cause of economic progress" [Han39].

2.2 Assumption II

The second assumption made is that the production of different economic structures has different levels of dependency on labor. Labor-intensive industries are usually traditional industries such as manufacturing and construction which are expected to benefit more from a larger labor force than those less labor-dependent ones. Whereas non-labor-intensive industries, such as finance, might generate higher output due to their greater leverage on the capital they own. To analyze the effect of the importance of the size of the labor force on production, we include an index named finance as an independent variable in our study, if it provides a more comprehensive explanation of the impact of labor on aggregate output.

3 Data

3.1 Data Collection

We collected our data from China's National Bureau of Statistics (NBS) which conducts a population census every ten years since 1950, which includes data on annual productivity surveys, labor force, average salary, etc. We obtained data of different features from the statistical department of the NBS, including production per capital for industries categorized into three levels, population and fatality rate divided by rural and urban regions, and participating labor force divided into rural and urban sections, all of which by province, with a time period ranging from 2001 to 2020. The data obtained included several null values in some of the years for the population section.

We put regions, their provinces, and the year when the data was collected into one index, and created a new column called college ratio which depicts the ratio of high school students having a college level or higher divided by the units of labor. We aim to find the education level by creating this column and see whether it is an indicator of Nominal GDP.

Description: df [10 × 7]

	Urbanization <dbl>	GDP <dbl>	Labor <dbl>	Finance <dbl>	Birth_rate <dbl>	College <dbl>	College_ratio <S3: AsIs>
黑龙江省-2016	0.6110309	11895.0	424.9	0.07268600	6.12	19.78	0.046552....
甘肃省-2014	0.4227578	6518.4	264.7	0.06891262	12.21	12.93	0.048847....
贵州省-2010	0.3380282	4519.0	224.3	0.05713654	13.96	9.93	0.044271....
湖南省-2002	NA	4151.5	NA	NA	11.56	15.69	NA
内蒙古自治区-2006	0.4865424	4161.8	NA	0.02724782	9.87	7.98	NA
北京市-2008	0.8492380	11813.1	570.3	0.13954800	8.17	15.61	0.027371....
青海省-2007	0.4003623	720.1	NA	0.03763366	14.93	1.11	NA
新疆维吾尔自治区-2008	0.3965275	4142.5	248.2	0.04482800	14.31	6.53	0.026309....
内蒙古自治区-2016	0.6338259	13789.3	293.2	0.05294685	9.03	12.19	0.041575....
广西壮族自治区-2019	0.5297069	21237.1	404.1	0.06919495	13.31	35.98	0.089037....

1–10 of 10 rows

Figure 1: Sample Rows of Data

3.2 Test for homocedasticity and autocorrelation

After cleaning our data, we first considered using Pooled OLS and Fixed Effects model. The residuals from the pooled OLS regression were used to test for Autocorrelation and Homoscedasticity, which are important assumptions in any regression analysis that must be satisfied to assure the validity of the results. We used the Breusch-Pagan test to test for Homoscedasticity. If the results of the test indicate that there is a constant variance for the residuals, it supports the assumption of homoscedasticity. Otherwise, the results would indicate possible heteroscedasticity, which can affect the validity of the regression results. Unfortunately, our Breusch-Pagan test results in a p-value low enough to reject our null hypothesis for Homoscedasticity, indicating Heteroscedasticity. For our testing of Autocorrelation, we used the Durbin-Watson test since it is commonly used for detecting autocorrelation in residuals from regression analysis. Typically, the test produces a result ranging between 0 and 4, with a value close to 2 indicating little or no autocorrelation in the residuals. In our case, the result was 0.51361, indicating a strong positive autocorrelation, thus Fixed Effect or Random Effect models are suitable for this dataset analysis. Based on the test results, we decided to use Panel Data Regression instead of Pooled OLS.

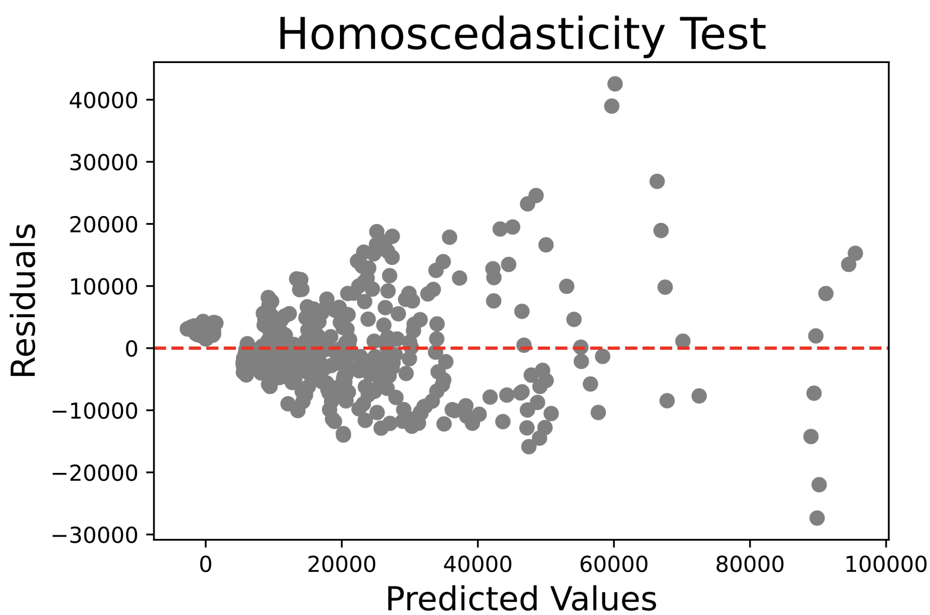

Figure 2: Homocedasticity Test

4 Methodology

In our research, we applied Panel Data Analysis for our interpretation of multifactor correlation and our prediction of future production in terms of Nominal GDP. We utilized the Fixed Effects Regression with the consideration of unobserved dependency of other independent variables on the observed one, holding those unobserved independent variables as constant overtime to minimize heterogeneity and ensure the existence of covariance between independent variables. We would consider Nominal GDP as our dependent variable, while the independent variables appeared in a step-by-step analysis as follows: labor force, finance, urbanization, birth rate, the other three were derived as individual effects while plugging each one as the independent variable for observation. After the regression for each independent variable was built with respect to the dependent variable, we performed a cross-variable comparison by reading their coefficients and testing their statistical significance by checking their p-values. Lastly, we interpreted the results and reached a conclusion. We would use log GDP as the dependent variable, and the independent variables included urbanization, log of labor units, finance, birth rate, and college ratio. The reason we used log of GDP and labor is because we wanted to minimize the variation of these two columns that had extremely high scales compared to others, and we could now see the percent change of these two variables rather than the oversized unit changes.

4.1 Pool Straight-up Regression

This method pools the data from multiple individuals or time periods and estimates a single regression equation for all observations. In contrast to the fixed-effects model and panel data regression, the straight-up pooled regression does not control for time-invariant variables and assumes that the relationship between the variables is the same for all individuals and across all time periods. The equation for a straight-up pooled regression can be written as follows:

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \epsilon_{i,t} \tag{1}$$

,where

1. $y_{i,t}$ is the dependent variable for individual i at time t $x_{i,t}$ is an independent variable for individual i at time t
2. β_0 is the intercept
3. β_1 is the coefficient for $x_{i,t}$
4. $\epsilon_{i,t}$ is the error term for individual i at time t

4.2 Fixed-Effects Model

Since there are unobserved time-invariant variables that may affect birth rate and income gap, we decided to use a fixed effect model. The model estimates the relationship between birth rate and income gap by removing the effect of those time-invariant variables, and focuses on the differences in the two variables across time for each individual. The fixed-effects model was estimated using the following equation:

$$Y_{it} = \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + \underbrace{\gamma_1 Z_{1i} + ... + \gamma_m Z_{mi}}_{=\alpha_i} + u_{it} \tag{2}$$

where i represents the individual (i.e., rural or urban area), t represents time. We first applied the region fixed effects regression, and then applied the region fixed effects but estimated in first differences regression.

4.3 Panel Data Regression

In addition to the fixed-effects model, a panel data regression was also estimated to control for both time-invariant and time-varying variables that may affect the relationship between birth rate and the income gap. The panel data regression allows for the analysis of the relationship between the two variables while controlling for both individual-specific variables (controlled by the fixed-effects model) and time-varying variables. The panel data regression was estimated using the following equation [Woo11]:

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \beta_2 z_i + \epsilon_{i,t} \tag{3}$$

,where

1. $y_{i,t}$ is the dependent variable for individual i at time t
2. $x_{i,t}$ is an independent variable for individual i at time t
3. z_i is a time-invariant independent variable for individual i

4. β_0 is the intercept

5. β_1 is the coefficient for $x_{i,t}$

6. β_2 is the coefficient for z_i

7. $\epsilon_{i,t}$ is the error term for individual i at time t

i represents the individual (i.e., rural or urban area), t represents time, and ϵ represents the error term. We first applied the region and time fixed effects regression, and then applied the regression with only time fixed as a comparison.

4.4 Statistical Analysis

All statistical analyses, Fixed Effect model, and Panel Data Regression were performed using Python and R Studio. The results of the estimation were used to analyze the relationship between birth rate and the income gap between rural and urban areas in China.

5 Discussions

5.1 Findings

Results for the straight-up pooled regression:

By looking at their p-values, we could see that all variables are statistically significant, and this is because there is large heterogeneity across regions and time. There exists omitted variable bias, although the results are significant.

Results for the region fixed effects regression:

This time, birth rate is becoming insignificant while others are still significant. Maybe birth rate has not an obvious relationship with GDP, but we still need to check it by adding the control for time.

```
##
## z test of coefficients:
##
##                 Estimate Std. Error z value  Pr(>|z|)
## (Intercept)   1.0131777  0.1586472  6.3864 1.699e-10 ***
## Urbanization  1.2064188  0.1709118  7.0587 1.680e-12 ***
## log(Labor)    1.0945919  0.0152898 71.5899 < 2.2e-16 ***
## Finance       2.8583162  0.6742181  4.2395 2.241e-05 ***
## Birth_rate    0.0583962  0.0048592 12.0176 < 2.2e-16 ***
## College_ratio 9.4148418  0.9473645  9.9379 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: Straight-up pooled regression

```
##
## z test of coefficients:
##
##                 Estimate Std. Error z value  Pr(>|z|)
## Urbanization  4.3924284  0.5463130  8.0401 8.974e-16 ***
## log(Labor)    0.8657813  0.1366146  6.3374 2.337e-10 ***
## Finance       3.0840124  1.5428487  1.9989 0.0456183 *
## Birth_rate    0.0050896  0.0104284  0.4880 0.6255156
## College_ratio 6.1050108  1.8010435  3.3897 0.0006997 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: Region fixed effects regression

Results for the region fixed effects regression estimated in first differences:

The first differences regression takes the delta value between years so that the time-invariant variable is eliminated, keeping the coefficients unchanged and record the delta change in variables. The assumptions under the first differences regression is the same with the normal fixed effects model. In this model, birth rate and college ratio are both having little association with Nominal GDP. This time, however, there are interesting findings. The coefficients for finance, birth rate and college ratio are turning negative,

```
##
## z test of coefficients:
##
##                 Estimate  Std. Error z value  Pr(>|z|)
## (Intercept)    0.08797485  0.00623837 14.1022 < 2.2e-16 ***
## Urbanization   1.40137073  0.31491194  4.4500 8.585e-06 ***
## log(Labor)     0.20719298  0.07261328  2.8534  0.004326 **
## Finance       -2.34247232  0.81831642 -2.8626  0.004202 **
## Birth_rate    -0.00083921  0.00238197 -0.3523  0.724598
## College_ratio -0.62878896  0.61235363 -1.0268  0.304496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: Region fixed effects regression estimated in first differences

by which we assumed to be positive because they should have a positive relationship with Nominal GDP.

Results for the region and time fixed effects regression:

```
##
## z test of coefficients:
##
##                Estimate Std. Error z value  Pr(>|z|)
## Urbanization   1.2277355  0.4484721  2.7376 0.0061890 **
## log(Labor)     0.5289553  0.1450987  3.6455 0.0002669 ***
## Finance       -1.5154153  0.8119288 -1.8664 0.0619800 .
## Birth_rate     0.0090604  0.0100527  0.9013 0.3674355
## College_ratio  3.2401107  1.6932715  1.9135 0.0556814 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 6: Region fixed effects regression estimated in first differences

This is the most comprehensive model since it takes into accounts both regional and time heterogeneity, but this time only urbanization and labor are significant. Finance still has a negative coefficient, but it can make sense because we get the conclusion that capital production is not the greatest indicator to Nominal GDP. There are other aspects in the industry that generate activity inside the economic system to boost the Nominal GDP.

Results for the time fixed effects regression:

```
##
## z test of coefficients:
##
##                Estimate Std. Error z value  Pr(>|z|)
## Urbanization   0.832742   0.370352  2.2485 0.0245433 *
## log(Labor)     1.081261   0.031863 33.9347 < 2.2e-16 ***
## Finance        0.335597   2.151472  0.1560 0.8760451
## Birth_rate     0.036157   0.010811  3.3443 0.0008249 ***
## College_ratio 5.384494   1.940591  2.7747 0.0055258 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7: Region fixed effects regression estimated in first differences

This time, birth rate is becoming super significant, which is a great jump from the fourth model. We should think about it because this time we are ignoring the regional heterogeneity, but if birth rate turns to be significant, the conclusion is that what matters to the determination of Nominal GDP is actually regional heterogeneity but not birth rate. Urbanization and labor force for all the five models are significant though, so these two indicators play vital roles in China's economic outputs.

5.2 Limitations

The limitations of our research come from two main sources: data and methodology.

With regards to data, our sample size is limited and this can have an impact on the generalizability of our findings. We should notice that the birth rate should have a delayed effect on production. For example, people born in the year 2000 will enter the workforce after receiving some education (normally it takes about 20 years or more, including 9 years of Compulsory Education and higher education), because of the limitation of our data size, we cannot include the lagged effect in our research. In addition, there have been difficulties in obtaining certain types of data, which could potentially affect the accuracy of our results. The labor data we used in our research is the labor force participation data in urban areas, because statistics regarding the rural workforce are not recorded regionally. We strongly suggest that any further research should include the total labor force. Finally, the reliability of the data must also be considered as it is possible that some data sources may not be completely accurate or up-to-date. Our research is done based on the data collected by China's National Bureau of Statistics. Due to the fact that the data we used are focusing on China's situation, our results could not be generalized to any other country. Also, we recommend further research to collect the data from other trustable databases or use primary data sources, to cross-valid our results.

In terms of methodology, there are some limitations that should be noted. Firstly, we may have not fully considered the interaction effects of variables in our analysis, which could potentially lead to biased results. Secondly, there is a potential for multicollinearity among variables, which could affect the precision and stability of our results. Finally, it is difficult to determine the effectiveness of change in birth policies and the lagged effect related to the change in birth rate. This is because changes in birth policies may take time to take effect and there may be other factors affecting the birth rate, making it challenging to isolate the effect of the policies.

In conclusion, the limitations of our research must be taken into consideration when interpreting our findings. While we have attempted to minimize these limitations, it is important to acknowledge that they do exist and to consider them when drawing conclusions from our results.

5.3 Conclusion

Despite being complicated and nuanced, the relationship between population and economic growth has received close attention from researchers and policymakers. In the case of China, it is widely accepted that the country's rapid economic development is due to the vast labor force that results from its enormous population. Based on this context, this study aimed to investigate the effect of China's birth rate on aggregate production, to examine the contribution of the labor force to economic growth, and evaluate the influence of the Family Planning Policy on economic development.

The relationship between the labor force size and China's overall output is investigated using panel data regression and fixed effect models. The information utilized in this study was gathered over a 20-year period (2000-2020) in order to provide a thorough understanding of the relationship between China's population expansion and its economic development. In order to account for additional variables that may have an impact on the relationship, many others were added to the labor force size in the analysis. These factors include our finance index, human capital, and economic openness indices.

Our analysis reveals that the size of the labor force does play a critical role in maintaining China's economic growth, the same as our initial hypothesis. The results of this study suggest that a larger labor force leads to an increased total production, which in turn leads to an increase in labor demand and promotes population growth, verifying Adam Smith's statements. However, the study also shows that the relationship between the labor force and economic growth is not a case that holds true permanently, as it depends on the structure of the nation's economy. Labor-intensive industries, such as manufacturing and construction, are likely to benefit more from a larger labor force than less labor-intensive ones, such as financial firms or technology companies. In conclusion, this study shows that the labor force is an important factor in sustaining China's economic growth. Our findings also suggest that population policies, such as China's Family Planning Policies, can have a considerable impact on the size of labor force and also economic growth. We believe that more research is needed to fully comprehend the complex relationship between population and economic growth, as well as to investigate any long-term impact of population control policies on economy.

5.4 Recommendation

Based on our findings, we would like to suggest some potential directions for future research. First, it would certainly be more beneficial to obtain a larger sample size and consider a more diverse range of industries to gain a deeper understanding of the impact of the birth rate on production. Also, it would be worth considering taking other variables into account for research, such as technological development, which should provide a better understanding of the relationship between birth rate and production. We hope these recommendations will contribute to a more comprehensive understanding how birth rate affects economic growth and provide insights for future policies such as China's Family Planning Policy.

References

[Han39] Alvin H Hansen. Economic progress and declining population growth. *The American economic review*, 29(1):1–15, 1939.

[Woo11] Jeffrey M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT Press, 2011.

[ZH16] Yi Zeng and Therese Hesketh. The effects of china's universal two-child policy. *The Lancet*, 388(10054):1930–1938, 2016.