4-2024

Three Essays on Fertility Rates in China

Lichun Cai
Western Michigan University

THREE ESSAYS ON FERTILITY RATES IN CHINA

Lichun Cai, Ph.D.

Western Michigan University, 2024

China is currently facing the issue of low fertility rates. In three chapters, I discuss the reasons for the low fertility rate in China from three different microeconomics perspectives, by using econometric methods including machine learning tools.

Chapter One considers the effects of income on fertility rate. There is a widely accepted view that the improvement of the education level has led to a decline in fertility. I critique this view and propose a new perspective: the childbirth choice of a family member is not only related to the education level, but also to the differences in education level between family members. Differences in education level among family members lead to a clear family division of labor, resulting in increased family income and fertility rates, thereby improving overall family welfare. My empirical analysis finds that differences in education level among family members do indeed have a significant positive impact on family fertility rates. Besides, such a positive impact is related to the gender differences in families and the overall difference in education levels between families.

Chapter Two considers the effects of income on fertility rate. The correlation between income and fertility rate has always been controversial. Most research shows that income and fertility rates are simply negatively related, while others indicate a more complicated

relationship. I first try to identify two different types of income, each affecting fertility rates differently. I also consider the income from assets like real estate, and the other variables related to income such as “Hukou” and job type. My finding is that the negative correlation between income and fertility rate indeed exists, but it is only in the sample of people who are engaged in agriculture jobs. There is no statistical correlation between income and fertility rate for those engaged in other jobs. Compared to income, there is a very clear positive correlation between house size and the number of children.

Chapter Three considers the effects of social activities on fertility rate. I discuss the distinction between social and non-social activities and their potential impact on fertility rates. An individual’s time opportunity costs are limited, and they must be allocated between social and non-social activities. The growth of non-social activities in modern society, driven by new entertainment products and lifestyles, has reduced the time for social activities, leading to a decline in fertility rates. In the empirical section, I attempt to use machine learning methods to help identify effective variables from a range of factors, specifically including two models using shrinkage methods like Ridge and Lasso. I find that certain activities, such as using the Internet, have a negative impact on fertility rates, especially some activities like having a cell phone or surfing the internet have a significant negative impact on the number of children.

THREE ESSAYS ON FERTILITY RATES IN CHINA

by

Lichun Cai

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Applied Economics
Western Michigan University
April 2024

Doctoral Committee:

Wei-Chiao Huang, Ph.D., Chair
Jean Kimmel, Ph.D.
Kevin H. Lee, Ph.D.

ACKNOWLEDGEMENTS

As I reflect on the journey of my doctoral dissertation, I am filled with immense gratitude for the exceptional individuals whose support and guidance have been invaluable. It is with heartfelt appreciation that I acknowledge their contributions.

My deepest thanks go to Dr. Wei-Chiao Huang, my dissertation committee chair, whose expert guidance and mentorship were pivotal in shaping my research and academic journey. His wisdom and support have been indispensable. Dr. Jean Kimmel's expertise and critical perspectives greatly enriched my work, and for this, I am truly thankful. I am profoundly grateful to Dr. Kevin Lee whose insightful suggestions significantly enhanced my research, providing me with a deeper understanding of complex statistical concepts.

A special tribute is paid to Dr. Mark Wheeler. His teachings and kindness have left a lasting impact on my life and career. Though he is no longer with us, his memory and wisdom continue to inspire me. Dr. Eskander Alvi deserves special mention for his extensive advice on conducting research. His suggestions were crucial in navigating the intricacies of my study, and I am immensely thankful for his guidance.

To all these remarkable individuals, I owe a profound debt of gratitude to my parents. Their endless love, support, and sacrifices have been the foundation of my achievements. Their belief in me has been my greatest motivation.

Lichun Cai

TABLE OF CONTENTS

Table of Contents—Continued

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

China's contemporary demographic landscape is characterized by notably low fertility rates, posing significant socio-economic challenges. This paper delves into the multifaceted reasons behind this trend, scrutinizing it through three distinct microeconomic lenses. The investigation is anchored in robust econometric methodologies, including advanced machine learning tools, to provide a nuanced understanding of the factors influencing fertility rates in China.

The first chapter scrutinizes the relationship between education levels and fertility rates, challenging the conventional notion that higher education levels invariably lead to reduced fertility. It introduces a novel perspective by emphasizing the role of educational disparities within families. The analysis reveals that variations in education levels among family members contribute to distinct family labor divisions, which paradoxically might enhance both family income and fertility rates, thus improving overall family welfare. This effect is further nuanced by gender differences within the family unit and variations in educational levels across different families.

In the second chapter, the focus shifts to the intricate and often controversial correlation between income and fertility rates. This segment of the paper dissects income into two categories: one relying on human capital accumulation and the other stemming from non-human capital sources like land and capital. A key finding here is the differentiated impact of these income types on fertility rates, with a distinct negative correlation evident only among those engaged in agricultural jobs. Interestingly, this essay also uncovers a strong positive correlation between house size and the number of children, indicating a more complex relationship between economic factors and fertility decisions.

The third chapter explores the impact of social versus non-social activities on fertility rates. It posits that the rising prevalence of non-social activities in modern society, driven by

new forms of entertainment and changing lifestyles, has encroached upon the time allocated for social activities, subsequently contributing to a decline in fertility rates. This section employs machine learning methods to sift through various factors, concluding that certain activities, such as internet usage, have a discernible negative impact on fertility rates.

The paper aims to provide a comprehensive and empirically grounded understanding of the dynamics shaping fertility rates in China, offering insights that are crucial for policy-making and future research in the field.

I. The Effect of Education Differences on Fertility Rate

1. Introduction

The reduction in fertility has become a global economic phenomenon. The fertility rate of most countries has shown a downward trend after World War II. According to the United Nations data, the global fertility rate dropped from 5.0 to 2.4 from 1960 to 2018. Europe registers one of the most significant declines in the global fertility rate (Lutz, 2005). The most significant decline is in Italy and Spain (Perez and Bacci, 1992). Their fertility rate dropped from 6 children/women in the early 1960s to 1.4 children/women in 2018. The fertility rate of developing countries is also declining, although the fertility rates of these countries are still relatively high. For example, China's fertility rate fell from 5.8 children/woman in the 1970s to 1.6 children/woman in 2018 (Jiang, Yang, Li and Feldman).

Decrease in fertility will have a significant impact on the demographic structure. Due to the increase in population and decline in fertility, the demographic structure of many countries has changed, and the aging population has become a serious problem. As the fertility rate decreases, the labor market may become more tense because the supply of labor will decrease. In addition, countries with low fertility rates may face shortage of labor and lack of domestic consumption (Robey, Rutstein and Morris, 1993).

Among all the theories of the decline in fertility rates, the most widespread one is education as a decisive factor (Martin, 1995). The improvement of education level leading to a decline in fertility has been widely confirmed. However, it is also unlikely that the improvement of education level is the only factor of the declining fertility rate. This paper proposes a new perspective to interpret the role of education. We will analyze the role of the difference of the education levels between family members on fertility rate.

There are four sections in this paper. The first section introduces the motivation of this research, first presenting the role of education and criticism of it, and then we introduce

the concept of difference of education level from the perspective of family labor division, and finally briefly discuss the education culture in China and its impact. In the second section, we provide a description of the sample data. The third section is empirical analysis, where we start from the baseline model and further analyze the impact of education difference, effects of gender difference in education and effects of family difference in education. The fourth section is the conclusion.

2. Motivations

2.1 Education as an effective influential role and criticism of it

The impact of education on fertility rate is roughly manifested in three aspects: the increase in opportunity costs, the decreased demand for children as labor, and improvement of women's status.

With the improvement of education level, people attach more attention to self-development and career goals, which means people are more willing to invest time and resources for learning and career development. If people decide to have children, they must give up one or more of these opportunities. Therefore, people may be more inclined to choose career or self-development over having children (Bongaarts, 2001). In addition, raising the education level means that the years people receive education must be extended, which inevitably conflicts or delays the time to have kids, and the time for child-caring, thereby reducing the desired number of children.

Some of the latest studies revealed some other possible effects, such as the change of lifestyle. These views emphasize changes in cultural and social values. With the improvement of education, some people are more inclined to pay attention to self-development and pursue personal achievements, rather than gaining social status by fertility (Mason, 1997). The transformation of these cultures and social values can also lead to a decline in fertility.

Modern education makes people give up the traditional multi-child lifestyle. On the other hand, modern industrialized society provides a variety of entertainment products. The increase of modern entertainment products gives people more options to spend their time. This may lead to people spending more time enjoying leisure activities rather than investing time in busy work and family life. As a result, people may give less thought to fertility issues or have less time for child caring.

In traditional society, children are often regarded as part of labor resources. In this case, the fertility rate is usually high, because families need more labor to support family livelihood. However, with the improvement of education level and the improvement of social-economic conditions, this situation has begun to change. With the improvement of education level, parents have begun to pay more attention to their children's education and training. They hope that their children will receive good education and get better career opportunities and better quality of life (Bongaarts, 2003).

In addition, labor demand in modern society is changing. With the advancement of technology and the changes in the industrial structure, more and more work requires high skills and high-quality labor, rather than relying on a large number of low-skilled labor. This makes parents more inclined to provide children with higher quality education so that they can adapt to the labor market that changes in the future (Hoffman, Thornton and Manis, 1978).

More empirical evidence shows that the impact from the education level of women on fertility is greater than the education level of man. This is not surprising because fertility and parenting are generally more women's work rather than men. The improvement of education of women makes them more likely to pursue economic independent career development and self-realization, rather than focusing on family fertility, resulting in the improvement of the position of women in society, thereby reducing the dependence on marriage and children, and

reducing the demand of having more children (Galor and Weil, 1993). In addition, women who have been educated are more knowledgeable to use contraceptive measures.

However, the role of education is not completely uncontroversial. The extension of women's education time will indeed occupy women childbearing time, but women have a long childbearing window in their life. Unless women spend so much time receiving education that it takes up the childbearing window in their life, women will always have time to have children. With the help of medical technology, women have more methods to extend childbearing time (Rindfuss and Brewster, 1996). In 2018, the world total fertility rate was 2.4, which means that a woman has an average of 2.4 children in her life. Considering the time spent in a woman's life-giving birth to 2.4 children, women are unlikely to devote all the remaining time to education. So, education is unlikely to be a single factor affecting fertility.

Another problem is the uncertainty brought by income. Higher education usually leads to higher income levels. In a fully marketized economy, a family can choose to purchase services from the market to invest human capital in children, instead of parents raising children by spending their own time (Furtado and Hock, 2008). Unless we can prove that the direct contact between parents and children is an irreplaceable factor in the childcare process, most parental childcare behaviors can be substituted with purchasing services from the market. The reduction in childcare time caused by the extension of parents' own education time can be offset by the services purchased by parents with higher income, if parents' education can bring them enough income to purchase those services.

The negative relationship between education and fertility is an empirical conclusion derived from post-war data, but if we examine more historical periods, we will be surprised to find that education is not negatively related to fertility in all historical periods (Bacci, 2012). In many historical periods, there was no statistically significant correlation between education and fertility. These periods often have some similar characteristics, such as social

crisis, economic recession and political instability. In other words, the theory of education as a critical factor in fertility selection is likely an empirical result derived from specific samples and periods, rather than a universally applicable theory.

2.2 Family division and education level difference

In this study, we first assume that the classic theory is established, that is, there is a negative relationship between the education level of family members and the fertility rate. On this basis, we put forward a new assumption: the fertility rate is not only related to the education level of family members, but also positively correlated with the difference of education level between family members.

In microeconomics, there is a theoretical framework called comparative advantage. This principle means that under free trade conditions, various countries or regions would be producing goods or services that they are relatively good at, to obtain greater benefits from trade. The basic idea of comparative advantage is that even if one country is more efficient than the other country in producing all products, there are still potential benefits of cooperation between the two countries (Costinot, 2009). Because even if one country is better in the production of all commodities than another country, two countries will have different opportunities costs, which means that they will be more efficient in producing certain products than to produce other products. Through the production of products that they are relatively good at producing, countries can use comparative advantages to achieve higher efficiency production and gain greater benefits from trade.

In labor economics, we have a concept to describe the difference between family roles caused by comparative advantages among different family members, that is the family division of labor. Family division of labor refers to the way in which household tasks, childcare, and other family responsibilities are distributed among the members of a family.

This allocation of duties is often based on traditional gender roles, with men taking on more economic and external work responsibilities and women taking on more domestic chores and childcare duties (Kroska, 1997). An equitable family division of labor contributes to the overall well-being of family members by allowing them to balance their work and personal lives more effectively. This balance can lead to improved mental and physical health, stronger family relationships, and greater satisfaction with work and family roles.

However, the family division of labor is changing to be more balanced and equal between the genders. This change is a result of various factors. Traditional gender roles, which often assigned men the role of breadwinners and women the role of caregivers, are being reevaluated and challenged. Many people now recognize that these roles can be limiting and that both men and women can contribute to family life in diverse ways. This shift in attitudes encourages a more equal sharing of responsibilities within families (Coltrane and Shih, 2009).

As we mentioned before in discussing the role of education in low fertility rate, as women gain access to education and career opportunities, they are increasingly able to participate in the labor force and contribute financially to their households. This economic independence encourages a more balanced division of labor, as women are no longer solely responsible for domestic tasks.

On the one hand, as gender norms evolve, men are becoming more involved in domestic work. This includes tasks such as cooking, cleaning, and child-caring, which were traditionally considered “women’s work” (Ross, 1987). This increased involvement helps to balance the division of labor within families and supports gender equality.

Even though there is a move towards more equality in sharing work between men and women, it is still useful to recognize that some tasks are divided by family roles and this can sometimes be helpful. One reason for this is the differences in education among family

members, which can result in a more efficient allocation of tasks within the household based on comparative advantages.

When family members have different levels of education or specialized skills, it can be more efficient for them to divide tasks based on their respective strengths (Mederer, 1993). This concept of comparative advantage suggests that by specializing in what they do best, family members can optimize their productivity and contribute more effectively to the overall well-being of the family.

In a family, members with a higher level of education are more likely to increase family income because they usually have stronger professional skills and knowledge backgrounds, and can gain better achievements in occupations and careers, thereby bringing more to the family income and resources. At the same time, family members with a lower education level may pay more attention for childcare although they can also bring income to the family. Therefore, family members should use their comparative advantages to achieve the best allocation of resources and maximize economic benefits, thereby promoting the stability and sustainable development of the family.

We establish a model that includes the education level of various family members and the difference of education level of these family members in the empirical section, thereby verifying the specific relationship between these variables. In the next section, we will first describe the data used, and then give the results of empirical analysis.

2.3 Chinese education culture and its impact

China’s unique history and culture may impact the fertility strategies of Chinese families in ways that result in empirical outcomes diverging from the predictions of classical microeconomics.

The Imperial Examination system was an ancient Chinese method of selecting

officials, primarily based on exam results. The first Imperial Examination took place in 621 A.D., and the last was held in 1905 A.D.. A system enduring for more than a thousand years is sufficient to alter a nation's culture and forge a unique collective psyche (Chen, Kung and Ma, 2020).

For families, having a child to pass the examination and become an official symbolizes the highest honor and social status. To increase the chances of a child's successful passing of the exam, families might invest more resources in education. This could result in families choosing to have fewer children, ensuring that adequate educational resources could be given to each child. Such a strategy is consistent with the modern quantity-quality trade-off theory, which means that parents always make choices between having more children and having fewer children with higher investment for each one (Shiue, 2013).

There is another view indicating that quantity-quality selection strategy may not be applicable to the imperial examination system. One reason is that the privileges brought by the imperial examination, such as the tax-free qualification, can be shared by all family members. As long as one child passes the exam and becomes an official, the reward is very impressive for the whole family. Therefore, it is an appealing strategy that parents sacrifice their short-term living standards in order to obtain more children with higher education levels (Li and Wang, 2022).

The Imperial Examination system had a profound influence on China's education and fertility culture historically. Although modern educational systems and fertility strategies have changed significantly, the impact of the Imperial Examination system can still be observed in many aspects.

Although the Imperial Examination system was in place for such a long period of time, it is still hard to imagine that a medieval style system would influence the mindset of modern

Chinese people. However, if we examine the history of modern China, it is not surprising that Chinese families place a high value on education investment. China experienced severe internal conflicts from the 1960s to the 1970s, one result of which was the closure of a large number of schools nationwide, leading to a decline in education level for nearly an entire generation (Giles, Park and Wang, 2019).

Therefore, when this ancient country decided to step into economic transformation in the 1980s, it found such a lack of skilled labor that it could hardly support rapid economic development. The shortage of high-skilled labor had greatly increased the returns of the educated people. In 1977, the Chinese government restored the College Entrance Exam system, and universities began to admit students through this exam again as they did before the internal conflict period. Almost all university graduates from the late 1970s to the early 1980s achieved impressive returns on their personal career development, which led to a high expectation of investing in education nationwide (Wu and Devine, 2019).

On the other hand, starting from the 1980s, China entered a period of high economic growth, thus the Chinese people of that era had many choices. They could choose to have more children to increase labor in the family, or increase education investment for each of their children, or do both. All these choices heightened the demand for division of labor within the family, because more resources and parenting time are required whether it is to increase the number of children or to improve the investment for each of children (Li, 2003). The limited number of family members must enhance efficiency through division of labor, and the difference of education level between family members may be one of the reasons for family division of labor.

China's education culture makes it a unique sample for studying the division of labor within families and the difference in education levels among family members.

3. Data description

The data used in this paper is the Chinese General Social Survey (CGSS). This is a survey jointly launched by China Renmin University and other academic institutions over the country since 2003. The survey covers many aspects such as work, family and so forth. The survey provided us with random sampling cross-section data, including 783 variables and 12582 observation items. This study uses the latest data in 2017.

CGSS includes four parts A, C, D, and Z. Among them, part A is the core module, including 90 questions, covering all basic personal information. Section C includes 66 major issues about social activities. Part D contains 42 questions, mainly to further describe the social relationship of the respondents. The Z part contains 9 questions, which is supplemented by the personal information of the respondent.

We selected some relevant questions in the original survey to form a data set. There are some missing values in the dataset. All the observations associated with the missing values are dropped. For some variables, the value of the answer "I don't know" is set to 98 or 998, and the value of the answer "I refuse to answer the question" is set to 99 or 999. We drop all observations with values of 98, 99, 998, and 999 because these cannot be meaningful interpreted.

In the original survey, the variables of "education level", "spouse's education level", "father's education level" and "mother's education level" have a value of 14, which represents "other education level". For these 4 variables, we drop all observations with values equal to 14. After cleaning up the original data, the sample contains 8770 observations.

In Table 1.1, We further listed the data description of "education level" in detail. These descriptions are applicable to all the four related variables such as "respondent's education level", "spouse's education level", "father's education level" and "mother's education level". There are 13 values in the variable in total, which means different education

Table 1.1 Education Level

Education level	Education years	Obs.	Percentage	Description
1	0	931	0.106	No education
2	1	50	0.006	Basic knowledge
3	6	1979	0.226	Primary school
4	9	2742	0.313	Junior high school
5	12	104	0.012	Vocational high School
6	12	1024	0.117	Normal high school
7	12	393	0.045	Technical secondary school
8	12	48	0.005	Technical school
9	15	298	0.034	College (Admission through the National Unified Examination for College Admissions for Adults)
10	15	397	0.045	College (Admission through the college entrance examination)
11	16	207	0.024	Undergraduate ((Admission through the National Unified Examination for College Admissions for Adults))
12	16	499	0.057	Undergraduate (Admission through the college entrance examination)
13	19	98	0.011	Graduate and higher

Notes: Table gives education level for 8770 observations.

levels from “no education” to “graduate and above”. The value of this variable is still qualitative, and it can only roughly describe the differences in the education level respectively.

In Table 1.1, education level 1 represents “having received no formal school education”. Education level 5 represents vocational high school, which is a unique type of high school in China, where students primarily learn labor skills. It is generally believed that the education level of its graduates is lower than that of regular high school students. Both education level 9 and level 10 are college levels, differentiated by the entrance exams. In China, to be admitted to a college or university, one student must pass either the “college entrance examination” or the “national unified examination for college admissions for adults”, with the former being more difficult. The coursework and degrees received by students who pass these two different exams are also different. While the Chinese government declares that both degrees have equivalent value, the labor market generally perceives a clear difference between them. The same logic applies to education level 11 and 12.

Researchers (Guo, Huang and Zhang, 2019) discussed the Chinese education system, listed the education years of each education level, and the age of approaching the education

level. Through research we can transform the education level into the education years, making us able to quantitatively compare education differences. The transformation result is recorded in Table 1.1 “education years” column.

Table 1.2 shows the descriptive statistics of all variables. We observe that the mean value of children is 1.799, which means that each respondent has about 1.8 children on average. The respondent’s mean education years is 9.477, and the mean of spouse’s education years is 9.371, indicating that the average education years of sample data is around 9.4. In contrast, the average education years of respondents’ fathers is 4.627, and the average education years of respondents’ mothers is 2.994. In the sample data, the average education years of the respondents is obviously higher than the preceding generation.

The mean value of the variable “gender” in the sample data is 0.5214, which means that there are more women in the data set than men. In this variable, the value of “woman” is set to 1, and the value of the “man” is set 0.

We transformed the variable “birth year” in the original sample into a new variable called “age”. The variable “age” is equal to 2017, which is the year that survey was conducted, minus the variable “year of birth”. The mean of “age” is around 51, with a minimum value of 18 and a maximum value of 103, indicating that the sample data covers respondents of all ages.

The unit of variable “height” is centimeters. The unit of variable “weight” is “jin”.[1]

The variable “ethnicity” is treated as a dummy variable, which means now we divided the whole population in China into 2 ethnic groups as majority and minority. The majority is Han Chinese, and the minority is Chinese of other ethnic groups. So, in the variable we define the value “Han” as 1 and “non-Han” is defined as 0.

The variable “religion” is treated as a dummy. In the survey, each answer to the

[1] It is a traditional Chinese weight unit and equal to 1/2 kilogram.

Table 1.2 Descriptive Statistics

Variable	Obs.	Mean	Std.	Min	Max
No. of children	8770	1.799	1.169	0	32
Respondent's education	8770	9.477	4.997	0	19
Spouse's education	8770	9.371	5.063	0	19
SRS(sum of Respondent's education and Spouse's education)	8770	18.85	9.024	0	38
Father's education	8770	4.627	5.077	0	19
Mother's education	8770	2.994	4.419	0	19
Difference of respondent's education and spouse's education	8770	0.105	4.444	-19	16
Absolute value of the difference of respondent's education and spouse's education	8770	2.806	3.447	0	19
Difference of respondent's education and father's education	8770	4.85	5.187	-16	19
Absolute value of the difference of respondent's education and father's education	8770	5.618	4.344	0	19
Difference of respondent's education and mother's education	8770	6.482	4.913	-16	19
Absolute value of the difference of respondent's education and mother's education	8770	6.71	4.596	0	19
Gender	8770	0.521	0.500	0	1
Age	8770	51.1	14.331	18	103
Height	8770	163.9	7.916	120	190
Weight	8770	123.9	23.442	38	260
Ethnicity	8770	0.927	0.261	0	1
Religion	8770	0.899	0.302	0	1

Note: age is equal to 2017 minus the birth year.

question represents a particular religion. The proportion of atheists in the respondents is much higher than other religions, so we define the value of atheism as 0 and define the value of all kinds of other religions as 1.

4. Empirical analysis

4.1 Baseline model

We first set up the baseline model. The estimation method is OLS. The regression equation is:

$$Children_i = \beta_0 + \beta_1 REdu_i + D_0\beta_2 SEdu_i + D_1\beta_3 FEdu_i + D_2\beta_4 MEdu_i + X'\beta + \varepsilon_i \quad (1.1)$$

The dependent variable of the model is the number of children. β_0 is the interception term. $REdu_i$ represents the respondent's education level. $SEdu_i$ represents the spouse's education level. $FEdu_i$ represents the father's education level. $MEdu_i$represents the mother's

education level. β_1 、 β_2 、 β_3 and β_4 are the regression coefficients. ε_i is the regression disturbance. X' represents other control variables such as gender, ethnicity, religion and so forth.

The result is presented in Table 1.3. From the results of all the 5 specifications, we can find that the respondents' education level has a negative and significant effect on the number of children. This means that people will have less children with the improvement of their education level. The results of the second and fifth columns show that there is also a statistically negative effect between the spouse' education level and the number of children of the respondents, which is a result that meets our expectation.

Column 3 shows father's education has a negative impact on respondents' children number, and Column 4 shows mother's education has a negative impact on respondents' children number. However, Column 5 shows that if we put father's education and mother's education in the model simultaneously, they will not have a significant impact on respondents' children. From the above results, we can infer that not all the education level of family members has a significant effect on the number of children of respondents. The educational influence of the respondents themselves and their spouses is mostly important, but the influence from the respondent's parents is much smaller.

In Columns 1, 3, 4, "gender" has a significant negative effect on respondents' children number, which means that in the sample, females have fewer children than male. We will further discuss the role of gender when discussing the relative position of family members' education levels.

In all the results the variable "age" has a significant positive effect on the number of children. We cannot infer that the fertility rate declines over time from this result. One of the reasons why the elderly may have more children is possibly because they have lived for a long enough time, so they have invested a lot of time in parenting.

Table 1.3 The Baseline Result by OLS

	(1)	(2)	(3)	(4)	(5)
Respondent's education	-0.061***	-0.041***	-0.057***	-0.056***	-0.039***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Spouse's education		-0.036***			-0.034***
		(0.003)			(0.003)
Father's education			-0.01***		-0.001
			(0.003)		(0.003)
Mother's education				-0.016***	-0.01**
				(0.003)	(0.004)
Gender	-0.159***	-0.052	-0.151***	-0.149***	-0.051
	(0.03)	(0.031)	(0.03)	(0.03)	(0.031)
Age	0.024***	0.022***	0.023***	0.022***	0.021***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Height	-0.007***	-0.005*	-0.007**	-0.006**	-0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Weight	-0.002**	-0.001*	-0.002**	-0.002**	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ethnicity	-0.118**	-0.103*	-0.119**	-0.112**	-0.1*
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Religion	-0.207***	-0.195***	-0.206***	-0.204***	-0.194***
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Observations	8770	8770	8770	8770	8770

Note: Dependent variable is No.of children. Estimation by OLS. ***p<0.001, **p<0.01, *p<0.05

In addition, we cannot ignore the role of policies, such as the famous "one-child" policy. If the influence of the policy does exist, then the families affected by the policy should have fewer children than the families that have completed parenting before the policy is introduced. If we need to get the answer to this question, we can find some related panel data to determine the number of children at a certain time point, but it will be the content of another paper.

In all five columns, we find that "height" has a significant negative effect on the number of children. Similarly, "weight" has a significant negative effect on the number of children. Considering the relatively small values of all the five regression coefficients, we cannot determine the effect of respondent weight on the number of children.

In all five columns, "ethnicity" has a significant negative impact on the number of children, which means that in samples, the respondents from the ethnic minorities have more

children than the respondents with ethnic majority.[2]

On the other hand, the variable "religion" has a significant negative effect on the number of children, which means that if respondents have some religious belief, they are more likely to have more children.

As a part of a robustness test, we carried out the ordered probit regression. All the specifications used in ordered probit are the same as OLS regression, and the results are included in Table 1.4.

We can see that most of the results of the ordered probit regression are consistent with the results of OLS regression, so it enhances the credibility of the baseline model. The difference between the two regressions is that the father's education and mother's education have a higher statistical significance level in the result of ordered probit regression, which indicates that we may underestimate the impact from the education level of parents in OLS regression.

4.2 The difference of education level

To examine the influence of the difference of education among family members, we have created several new variables to express the education gap between respondents and other family members. The first variable "difference between respondents' education and spouse's education" is equal to the respondents' education years minus the spouse' education years. Similarly, "difference between respondents' education and father's education" is equal to the respondents' education years minus father's education years, and the "difference between respondents' education and mother's education" is equal to the respondents' education years minus mother's education years.

In the regression we take the absolute value of these variables, which means that we

[2] The Han Chinese, and on the other hand, ethnic minority in China means the ethnicity other than Han Chinese.

Table 1.4 The Baseline Result by Ordered Probit

	(1)	(2)	(3)	(4)	(5)
Respondent's education	-0.076***	-0.050***	-0.069***	-0.066***	-0.044***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Spouse's education		-0.050***			-0.046***
		(0.003)			(0.003)
Father's education			-0.021***		0.000
			(0.003)		(0.003)
Mother's education				-0.039***	-0.032***
				(0.003)	(0.004)
Gender	-0.198***	-0.045	-0.182***	-0.175***	-0.040
	(0.032)	(0.033)	(0.032)	(0.032)	(0.033)
Age	0.029***	0.027***	0.027***	0.025***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Height	-0.010***	-0.007**	-0.010***	-0.009***	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Weight	-0.001*	-0.001	-0.001	-0.001*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ethnicity	-0.176***	-0.158***	-0.178***	-0.163***	-0.150**
	(0.045)	(0.046)	(0.045)	(0.045)	(0.046)
Religion	-0.191***	-0.178***	-0.190***	-0.186***	-0.174***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Observations	8770	8770	8770	8770	8770

Note: Dependent variable is No.of children. Estimation by ordered probit. *p<0.05; **p<0.01; ***p<0.001

temporarily ignore the relative position of education level of one family member to another one, but only to consider the size of this difference. We will further discuss the relative position of the educational levels of family members in the analysis of the gender difference.

The regression equation is:

$$Children_i = \beta_0 + \beta_1 REdu_i + D_0\beta_2 SEdu_i + D_1\beta_3 FEdu_i + D_2\beta_4 MEdu_i + D_0\beta_5|dSEdu_i| + D_1\beta_6|dFEdu_i| + D_2\beta_7|dMEdu_i| + X'\beta + \varepsilon_i \quad (1.2)$$

Here, $dSEdu_i$ represents the difference between respondents' education and spouse's education, $dFEdu_i$ represents the difference between respondents' education and father's education, and $dMEdu_i$ represents the difference between respondents' education and mother's education. All the differences are absolute value. Other variables are the same as the baseline model. The estimation method is OLS and ordered probit, and the results are

Table 1.5 The Education Level Difference Result

	(1)	(2)
Respondent's education	-0.033***	-0.033***
	(0.007)	(0.007)
Spouse's education	-0.034***	-0.034***
	(0.003)	(0.003)
Father's education	-0.004	-0.004
	(0.004)	(0.004)
Mother's education	-0.009	-0.009**
	(0.007)	(0.007)
Difference between respondent's education and spouse's education	0.025***	0.025***
	(0.003)	(0.003)
Difference between respondent's education and father's education	-0.008	-0.008
	(0.005)	(0.005)
Difference between respondent's education and mother's education	0	0
	(0.008)	(0.008)
Gender	-0.052	-0.052
	(0.031)	(0.031)
Age	0.021***	0.021***
	(0.001)	(0.001)
Height	-0.004*	-0.004**
	(0.002)	(0.002)
Weight	-0.001*	-0.001
	(0.001)	(0.001)
Ethnicity	-0.101*	-0.101*
	(0.043)	(0.043)
Religion	-0.19***	-0.19***
	(0.037)	(0.037)
Observations	8770	8770

Note: Dependent variable is No.of children. (1) is estimated by OLS. (2) is estimated by ordered probit. All differences are absolute value. ***p<0.001, **p<0.01, *p<0.05

presented in Table 1.5.

From Table 1.5, we can see that the absolute difference between respondents' education and spouse's education have a statistically significant positive effect on the number of children. This shows that the greater the difference between the education level of respondents and spouses, the more children's respondents will have. On the other hand, we have not observed significance of the two variables, the difference between respondents' education and father's education and the difference between respondents' education and mother's education, which means that the difference in education level of respondents and parents may not have a significant effect on the number of children.

4.3 Gender difference

We have analyzed the impact of the absolute value of differences in education levels among family members. However, we have overlooked the relative position of education levels within the family, which practically means we ignored the relative gender differences in education. There are two possible scenarios regarding the difference in their education levels. In the first scenario, the husband has a higher education level than the wife, while in the second scenario, the wife has a higher education level than the husband. In the previous analysis, we took the absolute value of the difference in education levels, which means we treated these two scenarios as one.

Now we divide the original sample into two sub-samples in terms of the gender of the respondents to analyze gender difference. In the first sub-sample, the respondents are male, and in the second sub-sample, the respondents are female.

We establish the regression equations applicable to these two sub-samples:

$$Children_i = \beta_0 + \beta_1 REdu_i + \beta_2 D_0 dSEdu_i + X'\beta + \varepsilon_i \qquad (1.3)$$

The results are shown in Table 1.6. Columns 1 and 3 are the results of sub-sample 1, and columns 2 and 4 are the results of sub-sample 2. Columns 1 and 2 are the results of OLS regression, and columns 3 and 4 are the results of ordered probit regression. We can see that for all the four columns, the difference between respondent's education and spouse's education has a statistically significant positive effect on the number of children. This result is consistent with the analysis of the absolute value of education difference in the previous section.

We can see that the coefficient in Column 1 is greater than the coefficient in Column 2 (0.057 > 0.013). This result indicates that for male respondents, if his education year is one

Table 1.6 The Gender Difference

	(1)	(2)	(3)	(4)
Respondent's education	-0.071***	-0.076***	-0.091***	-0.103***
	(0.004)	(0.004)	(0.005)	(0.004)
Difference between respondent's education and spouse's education	0.057***	0.013**	0.072***	0.027***
	(0.004)	(0.004)	(0.005)	(0.005)
Age	0.024***	0.02***	0.028***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)
Height	-0.002	-0.007*	-0.005	-0.009**
	(0.003)	(0.003)	(0.003)	(0.003)
Weight	-0.002*	0	-0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Ethnicity	-0.042	-0.154*	-0.087	-0.222***
	(0.06)	(0.06)	(0.066)	(0.063)
Religion	-0.17***	-0.205***	-0.142*	-0.196***
	(0.057)	(0.049)	(0.063)	(0.051)
Observations	4197	4573	4197	4573

Note: Dependent variable is No.of children. Estimation by OLS. (1) (2) are estimated by OLS. (3) (4) are estimated by ordered probit. ***p<0.001, **p<0.01, *p<0.05

unit more than that of his spouse, he is likely to have 0.057 more children. However, for female respondents, if her education year is one unit more than that of her spouse, she is likely to have 0.013 more children. Similarly, the coefficient in Column 3 is greater than the coefficient in Column 4 (0.072 > 0.027), which is consistent with the OLS results. The above results show that although the education difference has a positive impact on the number of children for both male and female respondents, the impact from males is greater than that from females. These results indicate that there exists a gender difference in the impact of education difference on fertility.

In the above methods, we divide the original samples into multiple samples according to the gender. This method may cause some information loss problems. We have another method to directly test gender differences by adding a new interaction variable into the model. We can construct a new variable, namely gender multiplied by education difference and then add it into our regression equation. This will allow us to directly test whether the impact of the difference in education level between spouses on the number of children is dependent on gender difference. Now our regression equation is:

$$Children_i = \beta_0 + \beta_1 REdu_i + \beta_2 D_0 dSEdu_i + \beta_3 Gender * D_0 dSEdu_i + X'\beta + \varepsilon_i \quad (1.4)$$

The results are presented in Table 1.7. Column 1 shows the results of the OLS regression, and Column 2 shows the results of the ordered probit regression. In the results of column 1, we see that the interaction term has a statistically significant negative effect on the dependent variable, with a coefficient of -0.046. This means that for every unit increase in the variable "gender", the impact of the education difference on the number of children decreases by 0.046. As we have defined gender as a dummy variable, with the value of male as 0 and female as 1, the regression results indicate that there indeed exists a gender difference, and when the respondent is female, the impact of education difference on the number of children is smaller than male respondents.

4.4 Family difference

When we mentioned that there are 13 values of the variable "education", which corresponds to 13 different levels of degrees or 8 different education years in the data description, we inevitably picked up a new problem. The problem is that there may be similar differences in education levels or years between two families, or between two couples, but this difference may come from completely different educational backgrounds. For example, the husband in family A has not received any education, while the wife has a primary school education. Therefore, we can see from Table 1.1 that the difference in education years between them is 6 years. In family B, the husband has a primary school education, and the wife has a high school education, so the difference in their education years is also 6 years. In both families, A and B, the difference in education years is both 6 years, but the educational backgrounds of the husbands and wives are completely different. So, does the numerically

Table 1.7 The Gender Difference by Interaction Term

	(1)	(2)
Respondent's education	-0.074***	-0.097***
	(0.003)	(0.003)
Gender	-0.04	-0.031
	(0.031)	(0.033)
Difference between respondent's education and spouse's education	0.059***	0.076***
	(0.004)	(0.004)
Age	0.022***	0.026***
	(0.001)	(0.001)
Height	-0.005*	-0.007**
	(0.002)	(0.002)
Weight	-0.001	-0.001
	(0.001)	(0.001)
Ethnicity	-0.102*	-0.158***
	(0.043)	(0.046)
Religion	-0.192***	-0.175***
	(0.037)	(0.039)
Gender * Difference between respondent's education and spouse's education	-0.046***	-0.052***
	(0.005)	(0.006)
Observations	8770	8770

Note: Dependent variable is No.of children. (1) is estimated by OLS. (2) is estimated by probit. ***p<0.001, **p<0.01, *p<0.05

equal difference in education years have the same effect on the fertility decisions of these two families?

To address this issue, we first propose a new variable called the "Sum of Respondent's and Spouse's Education Levels", which is in short as SRS. The SRS is calculated by adding the respondent's education years and the spouse's education years. We use SRS to measure the overall education level of the household the respondent belongs to. From Table 1.2, we can see that the minimum value of SRS is 0, and the maximum value is 38. We can divide the sample data into smaller samples based on the range of SRS values to analyze the performance of educational level differences in these samples. We could divide the original sample into 39 sub-samples with different SRS values. For convenience, we only divide the original sample into 5 sub-samples. In each sub-sample, respondents have similar SRS levels. For example, in the first sub-sample, SRS is greater than 0 and less than 9, while in the second sub-sample, SRS is greater than 9 and less than 17, and so on.

Here our regression equation is:

$$Children_i = \beta_0 + \beta_1 REdu_i + \beta_2 SEdu_i + \beta_3 dSEdu_i + X'\beta + \varepsilon_i \qquad (1.5)$$

The results are included in Table 1.8 and Table 1.9. In the first column of Table 1.8, The difference between respondents' education and spouse's education has a statistically significant positive impact on the number of children, with a coefficient value of 0.325. This result tells us that for every unit increase in the difference in educational levels, the respondent will have an additional 0.273 children. In the second, third and 4th columns of Table 1.8, The difference between respondents' education and spouse's education still has a statistically significant positive impact on the number of children. In column 5 of Table 1.8, there is no longer a statistically significant relationship between the difference in education levels and the number of children. Table 1.9 also provides us with some similar results.

In summary, we find that in sub-samples where household members have a lower overall educational level, the difference in educational levels has a statistically significant positive impact on the number of children. However, in sub-samples with higher educational levels, the relationship between the difference in educational levels and the number of children is no longer statistically significant. This indicates that as the SRS increases, the impact of the difference in educational levels on the number of children gradually weakens. Therefore, it is entirely necessary to consider the influence of the overall educational level of household members when analyzing the difference in educational levels.

However, splitting original samples into smaller samples by using the sum of education levels as a grouping variable may result in some information loss. For instance, two families might have the same sum of education levels, but their internal distributions could be different. One family might have spouses with equal education levels, while in the other

Table 1.8 The Difference between Families by OLS

	(1) 0≤SRS<9	(2) 9≤SRS<17	(3) 17≤SRS<24	(4) 24≤SRS<31	(5) 31≤SRS≤38
Respondent's education	-0.317*	-0.028**	-0.051	-0.025*	-0.018
	(0.14)	(0.011)	(0.027)	(0.01)	(0.02)
Spouse's education	-0.294*	-0.014	-0.067*	-0.018	-0.042*
	(0.139)	(0.011)	(0.027)	(0.01)	(0.021)
Difference between respondent's education and spouse's education	0.325*	0.025***	0.053***	0.036***	0.015
	(0.14)	(0.005)	(0.012)	(0.007)	(0.019)
Gender	0.135	-0.095	-0.082	-0.015	-0.122*
	(0.126)	(0.06)	(0.064)	(0.049)	(0.053)
Age	0.047***	0.03***	0.013***	0.011***	0.018***
	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)
Height	0	-0.007	-0.009*	-0.001	-0.015***
	(0.007)	(0.004)	(0.005)	(0.003)	(0.004)
Weight	0.001	-0.002	-0.001	-0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Ethnicity	-0.218	-0.064	-0.207*	-0.168*	0.018
	(0.158)	(0.075)	(0.09)	(0.085)	(0.075)
Religion	-0.285*	-0.3***	-0.084	-0.052	0.016
	(0.132)	(0.066)	(0.078)	(0.065)	(0.072)
Observations	1080	2599	1995	1653	1443

Note: Dependent variable is No.of children. Estimation by OLS. ***p<0.001, **p<0.01, *p<0.05

family, one partner's education level might be far higher than the other ones. On the other hand, if the education level of the respondents follows some statistical distribution, then in families with a smaller sum of education levels, the difference in education levels between family members will also be smaller, which could possibly affect our interpretation of the regression results.

To resolve these issues, we consider treating the sum of education levels as a continuous variable rather than using it to group the sample. Then in our regression model, we include the difference in education levels between spouses, and an interaction term between the sum of education level and the difference, which means the product of the sum of education levels and the difference in education levels. This will allow us to directly test whether the impact of the difference in education levels between spouses on the number of children depends on the total education level of the family. An advantage of this method is

Table 1.9 The Difference between Families by Ordered Probit

	(1)	(2)	(3)	(4)	(5)
	0≤SRS<9	9≤SRS<17	17≤SRS<24	24≤SRS<31	31≤SRS≤38
Respondent's education	-0.263*	-0.026**	-0.042	-0.05*	-0.053
	(0.098)	(0.011)	(0.03)	(0.018)	(0.036)
Spouse's education	-0.254*	-0.022	-0.061*	-0.044	-0.069*
	(0.098)	(0.011)	(0.03)	(0.018)	(0.037)
Absolute value of the Difference between respondent's education and spouse's education	0.273*	0.021***	0.058***	0.063***	0.037
	(0.099)	(0.005)	(0.013)	(0.013)	(0.034)
Gender	0.127	-0.059	-0.042	-0.057	-0.252*
	(0.088)	(0.059)	(0.071)	(0.083)	(0.093)
Age	0.041***	0.032***	0.016***	0.019***	0.033***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Height	0.002	-0.004	-0.012*	-0.006	-0.029***
	(0.005)	(0.004)	(0.005)	(0.006)	(0.007)
Weight	0.001	-0.002	0	-0.001	0.002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Ethnicity	-0.201	-0.121	-0.268*	-0.302*	0.041
	(0.11)	(0.074)	(0.098)	(0.143)	(0.132)
Religion	-0.245*	-0.23***	-0.159	-0.038	0.006
	(0.092)	(0.065)	(0.085)	(0.11)	(0.128)
Observations	1080	2599	1995	1653	1443

Note: Dependent variable is No.of children. Estimation by ordered probit. ***p<0.001, **p<0.01, *p<0.05

that it utilizes all the information in the original sample without splitting it, thus handling the continuity of the data better and providing a more accurate estimate of the effect size and direction. Now our regression equation is:

$$Children_i = \beta_0 + \beta_1 REdu_i + \beta_2 SEdu_i + \beta_3 dSEdu_i + \beta_3 SRS_i * |dSEdu_i| + X'\beta + \varepsilon_i \qquad (1.6)$$

The results are shown in Table 1.10. We can see that the interaction term has a statistically significant effect on the dependent variable, with a coefficient of -0.001, which means that for each additional unit of the SRS variable, the effect of the education difference on the number of children will decrease by 0.001. Such a regression result indicates that there indeed exist differences among families. Families with higher overall levels of education

Table 1.10 The Difference between Families by Interaction Term

	(1)	(2)
Respondent's education	-0.037***	-0.067***
	(0.003)	(0.003)
Spouse's education	-0.032***	-0.067***
	(0.003)	(0.003)
Absolute value of the Difference between respondent's education and spouse's education	0.049***	0.032***
	(0.008)	(0.008)
Gender	-0.053	-0.047
	(0.031)	(0.033)
Age	0.021***	0.026***
	(0.001)	(0.001)
Height	-0.005*	-0.007**
	(0.002)	(0.002)
Weight	-0.001*	-0.001
	(0.001)	(0.001)
Ethnicity	-0.102*	-0.159***
	(0.043)	(0.046)
Religion	-0.187***	-0.172***
	(0.037)	(0.039)
SRS * Absolute value of the Difference between respondent's education and spouse's education	-0.001***	0
	(0)	(0)
Observations	8770	8770

Note: Dependent variable is No.of children. (1) is estimated by OLS. (2) is estimated by probit. ***p<0.001, **p<0.01, *p<0.05

show a smaller influence of the difference in education levels among their members on their number of children.

5. Conclusion

The research results first verified for the members of a family, the increase in education level has a significant negative effect on the number of children. Whether it is the respondent or the spouse of the respondents, their education level has a significant negative impact on the number of children. In contrast, the education level of the respondents' father and mother is much weaker.

If we only consider the absolute value of the difference in education level, the education difference between respondents and spouses is significantly positive related to the number of children. This supports our conjecture, that is, the "comparative advantage"

brought by the difference in education level between family members may exist.

The impact of the difference in education level between spouses on the number of children is related to the gender difference among the family members. When the respondent is female, the difference between her education level and her spouses has a smaller impact on the number of children she has.

When analyzing the differences between families, we find that the impact of educational level differences is more likely to be significant in households with relatively lower overall educational levels. As the overall educational level of household members increases, the influence of educational level differences on the number of children gradually weakens, until it is no longer statistically significant.

References

Bacci, Massimo Livi. 2012. "Low Fertility in Historical Perspective." *Population and Development Review*, 2012, Vol. 38, Population and Public Policy: Essays in Honor of Paul Demeny (2012), pp. 72-82.

Bongaarts, John. 2001. "Fertility and Reproductive Preferences in Post-Transitional Societies." *Population and Development Review*, 2001, Vol. 27, Supplement: Global Fertility Transition (2001), pp. 260-281.

Bongaarts, John. 2003. "Completing the Fertility Transition in the Developing World: The Role of Educational Differences and Fertility Preferences." *Population Studies*, Vol. 57, No. 3, 2003, pp. 321-336.

Chen, Ting, James Kai-sing Kung and Chicheng Ma. 2020. "Long Live Keju! The Persistent Effects of China's Civil Examination System." *The Economic Journal*, Volume 130, Issue 631, October 2020, Pages 2030–2064.

Coltrane, Scott and Kristy Y. Shih. 2009. "Gender and the Division of Labor." *Handbook of Gender Research in Psychology*, DOI 10.1007/978-1-4419-1467-5_17.

Costinot, Arnaud. 2009. "An Elementary Theory of Comparative Advantage." *Econometrica*, Volume77, Issue4, Pages 1165-1192.

Furtado, Delia and Heinrich Hock. 2008. "Immigrant Labor, Child-Care Services, and the Work-Fertility Trade-Off in the United States." *IZA Discussion Paper*, No. 3506.

Galor, Oded, and David N. Weil. 1993. "The Gender Gap, Fertility and Growth." *NBER Working Paper Series* #4550.

Giles, John, Albert Park and Meiyan Wang. 2019. "The Great Proletarian Cultural Revolution, Disruptions to Education, and the Returns to Schooling in Urban China." *Economic Development and Cultural Change*, Volume 68, Number 1.

Guo, Lijia, Jiashun Huang, and You Zhang. 2019. "Education Development in China: Education Return, Quality, and Equity." *Sustainability*, 2019, 11, 3750.

Hoffman, Lois Wladis, Arland Thornton, and Jean Denby Manis. 1978. "The Value of Children to Parents in the United States." *Journal of Population*, Summer, 1978, Vol. 1, No. 2 (Summer, 1978), pp. 91-131.

Jiang, Quanbao, Shucai Yang, Shuzhuo Li and Marcus W. Feldman. 2019. "The decline in China's fertility level: a decomposition analysis." *Journal of Biosocial Science*, Volume 51, Issue 6, November 2019, pp. 785 - 798.

Kroska, Amy. 1997. "The Division of Labor in the Home: A Review and Reconceptualization." *Social Psychology Quarterly*, Vol. 60, No. 4 (Dec. 1997), pp. 304-322.

Li, Haizheng. 2003. "Economic transition and returns to education in China." *Economics of Education Review*, Volume 22, Issue 3, June 2003, Pages 317-328.

Li, Logan (Geng) and Zhengwei Wang. 2022. "Understanding the long-term effects of Keju: The case of entrepreneurship in China." *Economics of Transitions and Institutional Change*, Volume30, Issue4, Pages 665-689.

Lutz, Wolfgang. 2005. "Fertility Rates and Future Population Trends: will Europe's Birth Rate Recover or Continue to Decline?" *International Journal of Andrology*, 29 (2006) 25–33.

Martin, Teresa Castro. 1995. "Women's Education and Fertility: Results from 26 Demographic and Health Surveys." *Studies in Family Planning*, Jul. - Aug. 1995, Vol. 26, No. 4 (Jul. - Aug. 1995), pp. 187-202.

Mason, Karen Oppenheim. 1997. "Explaining Fertility Transitions." *Demography*, Volume 34-Number 4, November 1997: 443-454.

Mederer, Helen J. 1993. "Division of Labor in Two-Earner Homes: Task Accomplishment versus Household Management as Critical Variables in Perceptions about Family Work." *Journal of Marriage and Family*, Feb. 1993, Vol. 55, No. 1 (Feb. 1993), pp. 133-145.

Perez, Margarita Delgado, and Massimo Livi Bacci. 1992. "Fertility in Italy and Spain: The Lowest in the World." *Family Planning Perspectives*, Jul. - Aug. 1992, Vol. 24, No. 4 (Jul. - Aug. 1992), pp. 162-167+171.

Rindfuss, Ronald R., and Karin L. Brewster. 1996. "Childrearing and Fertility." *Population and Development Review*, Vol. 22, Supplement: Fertility in the United States: New Patterns, New Theories (1996), pp. 258-289.

Robey, Bryant, Shea O. Rutstein and Leo Morris. 1993. "The Fertility Decline in Developing Countries." *Scientific American*, Vol. 269, No. 6, pp. 60-67.

Ross, Catherine E. 1987. "The Division of Labor at Home." *Social Forces*, Volume 65, Issue 3, March 1987, Pages 816–833.

Shiue, Carol H. 2013. "Human capital and fertility in Chinese clans before modern growth." NBER working paper 19661.

Wu, Bin and Nesta Devine. 2017. "Self-cultivation and the legitimation of power: Governing China through education." *Educational Philosophy and Theory*, Volume 50, 2018 - Issue 13.

II. Income and Fertility Rate, Evidence from China

1. Introduction

One of the most controversial topics in all factors affecting fertility rate is the impact of income. There are at least two contradictory views on the role of income. The first view is that income has a positive effect on fertility rate, which means that with the rise of individual or family income, family members tend to have more children (Ewer and Gardner, 1978). The second view is the opposite of the first one. It is believed that there is a negative relationship between income and fertility, which means that the fertility rate declines as the individuals or family income rises (Jones, Schoonbroodt and Tertilt, 2008). Two views have related research support, but there are also differences in research results.

The origin of the first point of view may still be the biggest supporter of this view, that is, the classic Malthusian theory. Malthusian population theory is a theory of population proposed by the 18th-century British economist Thomas Robert Malthus. The theory mainly holds that population growth will quickly surpass the growth of food supply, leading to food shortages and poverty (Galor and Weil, 2000). Specifically, Malthus believed that population growth is exponential, while food supply can only grow linearly. Therefore, when population growth exceeds food supply, food shortages and population decline will occur. Malthus pointed out that when population growth is too rapid, three factors will limit population growth: war, famine and disease. These factors will cause population decline and bring the population back to a level that is limited with food supply (Ohno,1996).

Malthusian population theory has attracted widespread attention and controversy from the moment it was proposed. Some people believed that he overestimated the food supply problem (Dyson, 2001), while others believed that he underestimated innovation and technological progress, which could increase food production and support population growth (Kögel and Prskawetz, 2001). Nevertheless, even the strictest critics cannot deny that

Malthusian theory provides at least a valuable theoretical framework where food supply has a positive effect on the population (Ehrlich and Kim, 2005).

The relationship between food supply and population growth in Malthusian population theory can give us some insights in the discussion of the impact of income on fertility rates. The sources of income are diverse besides food supply. People can earn income through employment for wages, entrepreneurship for profits, investments for returns, and so on. Higher income may provide better living conditions and resources, thus encouraging people to have more children.

On the other hand, the second view suggests a negative correlation between income and fertility rates. This view mainly comes from empirical studies that are based on post-war global economic data. Supporters of this view often emphasize education as a decisive factor (Kunz, 1965). The results of related empirical studies suggest that higher levels of education typically lead to higher income, and improvements in education levels have a negative impact on fertility rates (Martin, 1995). From these two findings, we can naturally infer a negative correlation between income and fertility rates.

As a further explanation of education decision theory, some researchers attribute the negative correlation between income and fertility rates to a higher opportunity cost of the people with higher income. When a family chooses to have a child, they need to consider the trade-offs involved, including spending less time on career advancement, education, personal interests, and other activities. Therefore, an increase in opportunity cost may lead families to be more inclined to give up the choice of having more children, thereby reducing the fertility rate (Jones, Schoonbroodt and Tertilt, 2008).

Except for two contradictory views, some other researchers believe that there is no simple correlation between income and fertility. In one of the empirical research projects, the researchers presume that income should have a positive impact on the size of the family, but

they found that this relationship is negative in the empirical test (Becker, 1960). Some studies have keenly observed that there are different forms of income. Different forms of income growth have a different role in fertility (Schultz, 2005).

It sheds new light on the impact of income on fertility rate for analyzing the impact of different types of income on fertility rate respectively. If different types of income have different or even opposite impacts on fertility rates, then mixing different types of income together naturally cannot give us accurate results. Therefore, before analyzing the impact of income in detail, we should first differentiate between different types of income and examine the impact of each type separately. On the other hand, we should also consider other variables that affect fertility in the model.

2. Fertility model-relevant variables

2.1 Income

Income can be divided into two types based on the factors that affect income growth. The first type relies more on the accumulation of human capital, including education and training, while the second type of income comes from factors outside of human capital accumulation, such as land and capital.

The first type of income, which relies more on the accumulation of human capital, typically means better job opportunities. Therefore, when people improve their skills and knowledge, they usually have better economic prospects and more stable income, which may encourage them to delay childbirth or reduce the number of children to focus on their career and personal development (Tournemaine and Luangaram, 2010). Therefore, the first type of income source is typically associated with lower fertility rates.

The second type of income comes more from factors outside of human capital accumulation, such as land and capital. When economic growth is driven by these factors,

people are more likely to increase the size of their families to better utilize these opportunities and resources (Kalemli-Ozcan, 2003). In addition, when people have more land and wealth, they are also more likely to rely on family labor to earn more income (Schultz, 1990). Therefore, the second type of income source is typically associated with higher fertility rates.

The boundary in income growth from these two types is not always clear. In most cases, income growth comes from multiple factors including labor, capital, land, and others. A simple way to handle this is to assume that income growth for rural residents is primarily due to factors that do not depend on human capital accumulation, especially in cases where most rural residents engage in traditional agriculture. Traditional agriculture has a lower requirement for human capital accumulation, so the common fertility strategy in rural areas is to increase the number of children to increase labor input, rather than by enhancing individual human capital accumulation (Kulu, 2011).

Correspondingly, the source of income growth of non-rural residents is relatively complicated. Non-rural residents may engage in work other than traditional agriculture, so their income growth comes from a variety of factors, one of which is the accumulation of human capital. In general, non-rural residents generally have a higher level of education than rural residents, so the average human capital accumulation level is also higher. The fertility strategy of non-rural residents may be quite complicated and affected by various factors (Lerch, 2019). In summary, distinguishing income by the factors that affect income growth is a valuable direction.

2.2 Housing

In recent years, China's real estate prices have continued to rise, and the proportion of real estate in family asset allocation has become higher and higher (Shen, 2012). In this study, we consider housing as a part of income. In China, there is no nationwide real estate tax,

which means people can hold real estate for a long term at a relatively low cost without paying additional taxes. If people are not required to pay property taxes, they would have a stronger incentive to hold larger properties, which means they are more likely to convert their liquid assets into real estate. Therefore, the size of one's home can be seen as a part of their income measurement (Zhu, Li and Guo, 2018).

There may be a positive correlation between the size of house and fertility rate in a family, because larger living space can provide better living conditions, making family members feel more comfortable and at ease. This comfortable and relaxed environment can improve the quality of life and life satisfaction of family members, thus promoting their sense of well-being and health status. In addition, larger living space can also provide more privacy and independent space, making family members more independent and autonomous, thereby increasing their sense of belonging and identity with the family. These factors can inspire family members' desires and confidence, thereby promoting an increase in fertility rates (Kulu and Vikat, 2007).

2.3 Hukou

There is a system called "hukou" in China. This system is a system that divides residents into different categories of identification such as rural resident, non-rural (or urban) resident, soldiers and other types (Liu, 2005). The Chinese government implements different policies to the residents holding different types of hukou.

For example, rural residents have the right to apply for a certain amount of land for cultivation and are also allowed to build houses on their assigned land. On the contrary, urban residents are not allowed to purchase rural land and houses. Therefore, the only way for urban residents to legally obtain houses is to buy houses from real estate companies or second-hand markets. Of course, these houses are in urban areas, not rural areas (Chen and

Fan, 2016).

Another example is the inequality in educational rights for different hukou holders. In China, the primary criterion for being admitted to a university is the result of the college entrance examination. However, the university admission quota is allocated by province, not distributed based on population. This leads to some provinces having more university admission quota than others, and students in these provinces are therefore more likely to be admitted (Fu and Ren, 2010).

If a student's hukou is in a province where the competition for admission is intense, even if they went to high school in a different province, they still have to participate in the college entrance exam according to the province of their hukou, facing fiercer competition. On the other hand, for the students with rural hukou, they might already be lagging in their studies at the stage of middle school or even elementary school due to a lack of educational resources. This impacts their performance in the college entrance examination and chances of further education to some extent.

The impact of the hukou system on the fertility rate is relatively complex. Firstly, the population policy for urban households is relatively strict and limits the number of births, while rural households are relatively more relaxed in the number of new births. This difference affects the fertility desires of different household populations. On the one hand, people with urban hukou are constrained by a relatively strict birth policy, resulting in lower fertility desires. On the other hand, people with rural hukou enjoy more relaxed policies and therefore have relatively higher fertility desires.

Secondly, urban householders have better access to education and career development opportunities, while those with rural householders have relatively weaker access. This means that urban householders are more likely to obtain higher income and better career development opportunities, which results in higher costs of having children, and thus their

fertility rate is lower. Rural householders have relatively less education and career development opportunities, and the cost of having children is relatively lower, so their fertility rate is higher.

Finally, the hukou system also affects the social security benefits for different hukou holders. Urban hukou holders enjoy relatively better social security benefits, including medical care, pension, and unemployment insurance. On the other hand, rural hukou holders have relatively lower social security benefits. This may cause rural families to be more inclined to have more children to obtain economic support from their children in their old age, thereby making up for the lack of social security (Cai, 2011).

When explaining the source of income, we assume that most rural residents are engaged in traditional agriculture, and this assumption is not necessarily all reliable as the economic evidence. One of the possibilities is that the holders of agricultural "hukou" do not necessarily engage in agricultural work. To resolve this problem, we can analyze work experience data to determine if a person is currently working in agriculture or not. Therefore, we must distinguish the hukou holding status and work of rural residents.

2.4 Education Spending

Investing more in children's education often leads to higher child-rearing costs, which negatively impact fertility rates (Borg, 1989). Parents may prefer to invest more resources in fewer children to ensure they receive good education and better future opportunities. This phenomenon has been observed in many countries, especially where the cost of education is relatively high. This strategy of parents is known as the quantity-quality trade-off, which describes the choice families make between the number of children and the investment for each of children (Rosenzweig and Wolpin, 1980).

The main point of this theory is that family resources are limited. Therefore, if

families choose to have more children, they might not be able to invest too many resources in each child. Conversely, if families decide to have fewer children, they would be able to invest more resources in each child, such as providing them with better education and health care.

The quantity-quality trade-off theory can help explain some phenomena, like why families might choose to have fewer children when education costs increase or income decreases. This theory is also used to explain the trend of declining global fertility rates, especially in societies where investment in education significantly affects children's future income.

However, this theory does not always apply completely. Some families may choose to have more children due to cultural, religious, or personal values, even if that means they cannot invest a lot of resources in each child. Similarly, certain policies or social programs may release the burden on families, allowing them to have more children while also providing sufficient resources for each child. For example, an increase in public spending on basic education is associated with significant reductions in household private tutoring spending in urban China (Cheng and Zhang, 2015).

2.5 Policy

Policy is an important factor when discussing the fertility rate in China. China's worldwide famous One-Child Policy began in late 1970s. At that time, the Chinese government was worried that the rapid population growth would lead to the problems of unsustainable development of society. This was the first time that the Chinese government had included controlling population growth in its national economic development plan. In 1982, One-Child Policy officially became a basic national policy of China, and in the same year in December, it was written into the Chinese constitution. In 1991, the Chinese government once again stated that it would implement the One-Child Policy to strictly

control population growth (Zhang, 2017).

The Chinese government has stated that this policy greatly reduced China's fertility rate, which helped China release the pressure of rapid population growth for a period. However, this policy also brought about some long-term social problems. Due to cultural preferences and policy pressures, many families chose to have boys rather than girls, leading to a severe gender imbalance. In addition, this policy also triggered the so-called "4-2-1" problem, which means four grandparents and two parents are to be taken care of by one child. This is a unhealthy family structure in demographic studies (Cameron, Erkal, Gangadharan and Meng, 2013).

After strictly implementing the policy for at least 30 years, this policy has dramatically changed. In 2016, the Chinese government changed the policy to allow all couples to have two children to combat issues such as aging population and gender imbalance. However, many families did not choose to have more children due to economic pressures and changes in lifestyle, so the birth rate did not rise as much as expected. So, in May 2021, the Chinese government changed the policy again to allow couples to have three children. However, it is still questionable that this new policy can effectively increase the low fertility rate in China.

2.6 Other Facts

Finally, we consider the impact from other important factors in the model. These factors include traditional factors such as education level of respondents, education level of parents, gender differences, ethnicity, religion, etc., which are included as control variables.

3. Data description

The data used in this paper is Chinese General Social Survey (CGSS). This survey has been jointly launched by China Renmin University and other academic institutions over the

country since 2003. The survey covers many aspects including work, family and so forth. The survey provided us with random sampling cross-sectional data, including 783 variables and 12582 observations. This study uses the latest data from 2017.

CGSS includes four sections A, C, D, and Z. Section A is the core module, including 90 variables, covering all basic personal information. Section C includes 66 major issues about social activities. Section D contains 42 variables, which is mainly to describe the social relationship of the respondents. Section Z contains 9 variables, which is the supplement of the personal information of the respondent.

We selected answers to some relevant questions in the original survey to form a data set. There are some missing values in the dataset. All the observations associated with the missing values are dropped. After removing all observations with missing values, we obtained a sample with 3,789 observations, which is significantly smaller than the sample we used in the first chapter, containing 8,770 observations. This is because, in this chapter, we included a variable “Spending on education”, which had a large number of missing values. When we cleaned up the missing values for this variable, the sample size was greatly reduced.

For some variables, the value of the answer “I don’t know” is set to 98 or 998, and the value of the answer “I refuse to answer the question” is set to 99 or 999. In addition, there are other types of answers, which are of obvious abnormal value. For these missing values or abnormal values, we first delete the missing or abnormal values from the variable, then take the average number or medium value of the remaining values, finally replace the original missing or abnormal value by the average or medium value.

After cleaning up the original data, Table 2.1 shows the mean value, standard deviation, maximum value and minimum value of all variables. The sample after cleaning up the data contains 3789 observations.

From the first line, we can know how many children the respondents have, which is

Table 2.1 Descriptive Statistics

	Mean	SD	Min	Max
Children	1.702	1.3915	0	32
Income	0.403	2.282	0	99.3
House size	1.134	0.958	0.05	15
Houses	1.111	0.674	0	11
Hukou	0.468	0.499	0	1
Job	3.027	1.797	1	6
Spending on education	0.271	4.548	0	100
Spending on education per child	0.180	3.203	0	100
Age	51.03	16.873	18	96
Education	5.163	3.239	1	13
Gender	0.536	0.499	0	1
Ethnicity	1.348	1.368	1	8
Religion	0.895	0.307	0	1
Height	1.639	0.080	1.3	1.9
Weight	1.227	0.236	0.45	2.4
Father's education	2.904	2.241	1	13
Mother's education	2.253	1.929	1	13
		N = 3789		

Note: The value of the variables "income", "spending on education" and "spending on education per child" is divided by 100000. The value of the variables "house size", "height" and "weight" is divided by 100. The unit of house size is square meter.

Data from CGSS

equal to the number of sons plus the number of daughters. The value is 1.702, which means that each of the respondents has slightly less than 2 children on average. There is a value of 32 in the variable "number of children", which is also the maximum value of this variable. This means that there is one respondent having 32 children in the sample, which is an extremely large size.

The average income is 40293 yuan. As shown in Figure 2.1, the probability distribution histogram tells us more about the distribution of income. The horizontal axis represents different income levels, and the vertical axis represents the number of people who have that income at that level. The tallest bar on the left side may indicate that more people have lower incomes, while the bars on the right side represent a decreasing number of people with higher income. This figure shows that in the sample of 3,789 respondents, over 1,000 people had an annual income of less than 5,000 yuan, indicating that the overall income level of the respondents was relatively low.

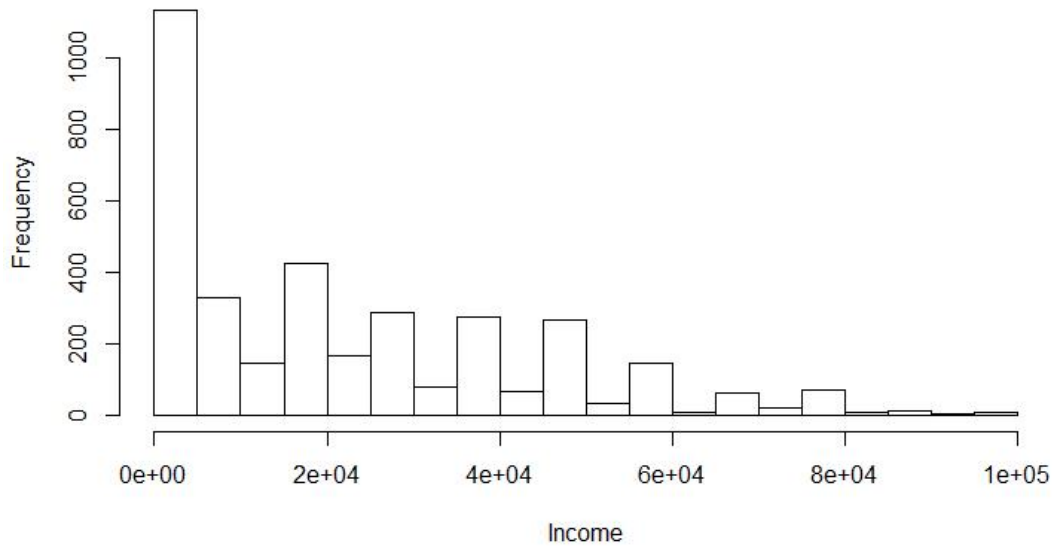

Figure 2.1 Histogram of Income

The variable "house size" refers to the actual usable area of a residential unit, including bedrooms, living rooms, kitchens, bathrooms, and other private spaces that can be directly used by the homeowner or tenant. It does not include public areas and non-usable spaces such as wall thickness. The unit of this variable is square meters.

Hukou is treated as a dummy variable, where the value of "rural" is defined as 0, and "non rural" is defined as 1. Job is also defined as a dummy variable, where non-agricultural job is defined as 0, and agricultural job is defined as 1.

The mean of the variable "spending on education" is 0.271, which means that on average, each respondent invests 27,100 yuan per year on their children for improving their children's education level. If we divide the mean of spending on education by the mean of income, the result is over 67%, which is an incredibly high proportion.[3] There is a possibility here that if the respondents also consider some other expenditures as education spending for

[3] In the original survey, the description of spending on education is "The proportion of children's education and training expenditure accounts for your family's total expenditure last year."

their children, they may overestimate the proportion of this expenditure in total family expenditure.[4]

The variable "spending on education per child" equals to the variable "spending on education" divided by the variable "number of children". From Table 2.1, we can see that there are values equal to 0 in the variable "number of children", meaning that some families currently have no children. Dividing any value of the variable "spending on education" by 0 would lead to results that are meaningless in algebraic terms. Therefore, these meaningless values are dropped from the variable. The mean, standard deviation, minimum and maximum values of "spending on education per child" seen in Table 2.1 are the results for this variable after dropping the meaningless values. The mean value of "spending on education per child" is 0.180, which is less than the mean of "spending on education" at 0.271, indicating that the investment in education per child is indeed lower than the total investment in children's education by families. This variable offers another perspective helping us to understand the impact of the net price per child on family fertility decisions.

We transformed the variable "birth year" in the original sample into a new variable called "age". The variable "age" is equal to 2017, which is the year the survey was conducted, minus the variable "year of birth". The mean of "age" is around 51, with a minimum value of 16.9 and a maximum value of 96, indicating that the sample data covers respondents of all ages.

Education has 13 discrete values. The values from low to high represent the 13 different levels of education levels from "no education" to "graduate students and above". The gap between different education levels may not be represented by an integer value, so this variable can only approximately describe the gap between education levels. There are 2

[4] The definition of spending on education in the original survey indicates that this variable may be endogenous, which means that it is a function of the number of children. There are no traditional exogenous variables about the net price of children such as tuition fee, textbook price, local childcare price and so forth. Including endogenous variables in the regression analysis may have an unpredictable effect on the results, so we look forward to solving this problem in future research.

values of gender where 0 represents men and 1 represents women.

The variables “ethnicity”, “religion”, “height”, “weight”, “father’s education level” and “mother’s education level” are control variables in the model. The eight values of “ethnicity” represent 8 major ethnic groups in China.

The variable “religion” is treated as a dummy. In the survey, each answer to the question represents a particular religion. The proportion of atheists in the respondents is higher than other religions, so we define the value of atheism as 0 and define the value of all other kinds of religions as 1.

The unit of height is centimeter. The unit of weight is a traditional unit “jin” in China, which means 1/2 kg. “Father’s education level” and “mother’s education level” are the same as “education level” in the form of data. They all divide the education into different levels by 13 values.

4. Empirical analysis

4.1 Baseline Model

Firstly, we established a baseline model. We took the number of children as the dependent variable, income as the main independent variable, and treated other variables as control variables. We magnified the units of some variables to make the results more intuitive and easy to explain. According to the descriptive statistics results of Table 2.1, the values of income and education spending were adjusted to 1/100,000 of the original values, the house size was adjusted to 1/100 of the original value, and height and weight were adjusted to 1/100 of the original values.

We use OLS as the first estimation method for the baseline model. An important assumption of the OLS method is that the error terms are independent and normally distributed with a constant variance. If these assumptions are not held, the least squares

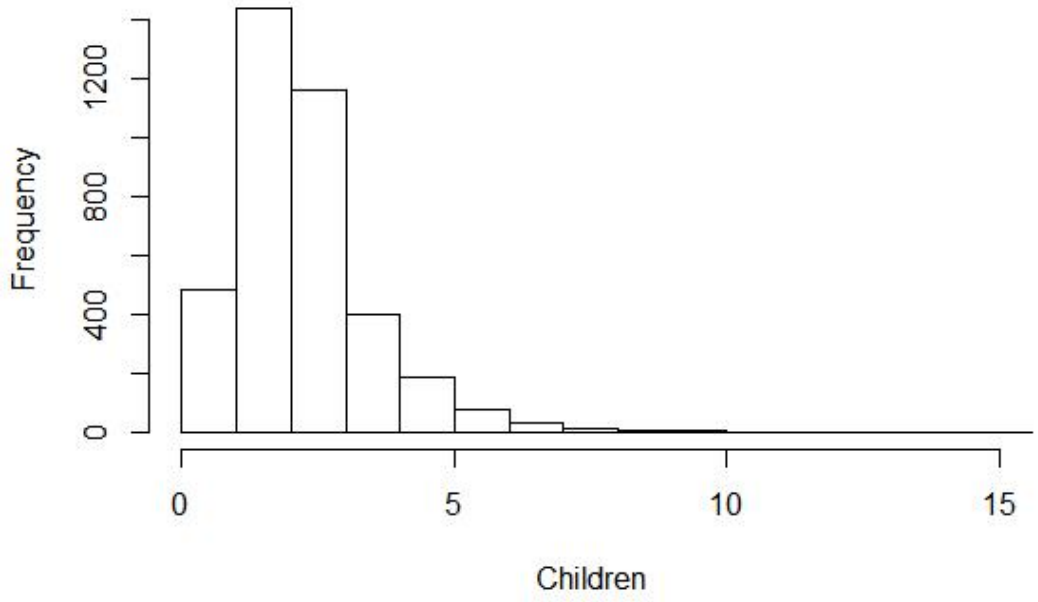

Figure 2.2 Histogram of Children

estimation could be biased. Therefore, we also applied other estimation methods besides OLS.

Our second estimation method for the baseline model is the Poisson regression, which is one of the count data models. As we can see in Figure 2.2, the dependent variable intuitively appears to follow a Poisson distribution. Poisson distribution is commonly used to describe the number of times that an event occurs in a fixed time, space, or area. This distribution has a characteristic that its mean and variance are equal. In our baseline model, the mean of the dependent variable is 1.702, and the variance is 1.392, which are reasonably close. Therefore, applying Poisson regression could be a suitable choice.

We use ordered probit regression as the third estimation method. Ordered probit regression is a robust regression method. This means that it can produce reliable results even if there are outliers or unreasonable values in the datasets. In contrast, other regression methods like linear regression or ordered logit regression might be more easily affected by outliers or unreasonable values, leading to unreliable results.

Table 2.2 Results of Baseline Model

	Dependent variable: number of children		
	(1)	(2)	(3)
Income	0.005	0.004	0.01
	(0.008)	(0.005)	(0.008)
House size	0.204***	0.095***	0.192***
	(0.02)	(0.01)	(0.019)
Houses	-0.007	0.003	0.041
	(0.027)	(0.019)	(0.027)
Hukou	-0.423***	-0.249***	-0.51***
	(0.048)	(0.035)	(0.048)
Job	0.214***	0.058	0.164***
	(0.05)	(0.033)	(0.049)
Spending on education	0.004	0.001	0.004
	(0.004)	(0.002)	(0.004)
Age	0.038***	0.022***	0.044***
	(0.001)	(0.001)	(0.002)
Education	-0.027**	-0.029***	-0.046***
	(0.008)	(0.006)	(0.008)
Gender	0.129**	0.071*	0.171***
	(0.05)	(0.033)	(0.049)
Ethnicity	0.004	0.003	0.013
	(0.014)	(0.009)	(0.013)
Religion	-0.219***	-0.101**	-0.204***
	(0.06)	(0.038)	(0.058)
Height	0.054	-0.126	-0.129
	(0.353)	(0.235)	(0.344)
Weight	-0.129	0.026	0.121
	(0.093)	(0.065)	(0.092)
Father's education	-0.013	-0.01	-0.018
	(0.012)	(0.01)	(0.012)
Mother's education	-0.007	-0.042***	-0.054***
	(0.014)	(0.012)	(0.015)
Observations	3789	3789	3789

Note: (1) is the estimation by OLS. (2) is the estimation by Poisson. (3) is the estimation by ordered probit. ***p<0.001, **p<0.01, *p<0.05

Data from CGSS

The results of the baseline model are in Table 2.2. Following the baseline model, we split the original sample into two smaller samples in terms of the hukou. In the first sample all the respondents have rural hukou, and in the second sample all the respondents have non-rural hukou. We applied the same three estimation methods to these two samples as we do in the baseline model, and the results are in Table 2.3. We then split the original sample in terms

Table 2.3 Results of Hukou Status

	Dependent variable: number of children					
	Rural hukou holders			Non-rural hukou holders		
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.003	0.002	0.004	0.039	0.046	0.111**
	(0.01)	(0.005)	(0.008)	(0.025)	(0.025)	(0.034)
House size	0.169***	0.077***	0.136***	0.263***	0.122***	0.299***
	(0.027)	(0.013)	(0.022)	(0.027)	(0.018)	(0.035)
Houses	-0.103*	-0.05	-0.042	0.051	0.048	0.09*
	(0.046)	(0.028)	(0.038)	(0.028)	(0.028)	(0.039)
Job	0.027	0.02	0.01	0.429***	0.176*	0.45***
	(0.069)	(0.037)	(0.056)	(0.093)	(0.076)	(0.119)
Spending on education	-0.004	-0.002	-0.003	0.016**	0.006	0.014*
	(0.006)	(0.003)	(0.005)	(0.005)	(0.003)	(0.006)
Age	0.044***	0.021***	0.045***	0.032***	0.023***	0.044***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Education	-0.058***	-0.055***	-0.089***	-0.015	-0.014	-0.023*
	(0.016)	(0.011)	(0.014)	(0.008)	(0.008)	(0.011)
Gender	0.184*	0.077	0.203**	0.075	0.062	0.129
	(0.076)	(0.041)	(0.063)	(0.057)	(0.059)	(0.078)
Ethnicity	-0.001	0.001	0.007	0.035	0.026	0.056*
	(0.018)	(0.009)	(0.015)	(0.021)	(0.02)	(0.029)
Religion	-0.279**	-0.117*	-0.277***	-0.158*	-0.075	-0.103
	(0.097)	(0.049)	(0.08)	(0.065)	(0.061)	(0.087)
Height	-0.073	-0.095	0.018	0.03	-0.111	-0.315
	(0.535)	(0.284)	(0.442)	(0.414)	(0.423)	(0.566)
Weight	-0.122	0.05	0.153	-0.091	-0.01	0.021
	(0.149)	(0.083)	(0.124)	(0.104)	(0.109)	(0.143)
Father's education	0.006	0.001	0.007	-0.022	-0.02	-0.036*
	(0.024)	(0.015)	(0.02)	(0.011)	(0.013)	(0.016)
Mother's education	-0.018	-0.039*	-0.052*	-0.019*	-0.044**	-0.078***
	(0.031)	(0.02)	(0.026)	(0.013)	(0.016)	(0.019)
Observations	2017	2017	2017	1772	1772	1772

Note: (1)(4) are the estimation by OLS. (2)(5) are the estimation by Poisson. (3)(6) are the estimation by ordered probit.
***p<0.001, **p<0.01, *p<0.05

Data from CGSS

of job, with one sample representing respondents engaged in agricultural work, and another sample representing those engaged in non-agricultural work. The results are in Table 2.4.

Since the sample used in Table 2.3 is divided according to hukou, the model corresponding to Table 2.3 has not included the independent variable of hukou compared to the baseline model. Similarly, the sample used in Table 2.4 is divided according to job, so the model corresponding to Table 2.4 has not included the independent variable of job compared to the baseline model, and the row of job has also been deleted in Table 2.4.

Table 2.4 Results of Job Sector

	Dependent variable: number of children					
	Agriculture job			Non agriculture job		
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.328**	0.114**	0.162*	0.003	0.002	0.009
	(0.12)	(0.04)	(0.082)	(0.006)	(0.005)	(0.008)
House size	0.245***	0.09***	0.145***	0.178***	0.096***	0.213***
	(0.048)	(0.017)	(0.032)	(0.018)	(0.014)	(0.023)
Houses	-0.154	-0.051	0	0.024	0.021	0.049
	(0.087)	(0.038)	(0.058)	(0.023)	(0.022)	(0.03)
Hukou	-0.334	-0.157*	-0.293**	-0.422***	-0.294***	-0.556***
	(0.173)	(0.074)	(0.118)	(0.041)	(0.04)	(0.055)
Spending on education	-0.01	-0.004	-0.008	0.01**	0.004	0.01
	(0.009)	(0.005)	(0.006)	(0.004)	(0.003)	(0.005)
Age	0.046***	0.02***	0.044***	0.034***	0.023***	0.044***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
Education	-0.093*	-0.035*	-0.04	-0.025***	-0.025***	-0.046***
	(0.037)	(0.016)	(0.025)	(0.007)	(0.007)	(0.009)
Gender	0.258*	0.113*	0.268***	0.065	0.041	0.116
	(0.118)	(0.049)	(0.08)	(0.047)	(0.047)	(0.062)
Ethnicity	0.011	0.005	0.021	0.005	-0.001	0.001
	(0.025)	(0.01)	(0.017)	(0.016)	(0.015)	(0.021)
Religion	-0.287	-0.101	-0.264**	-0.191***	-0.105*	-0.184**
	(0.155)	(0.06)	(0.104)	(0.054)	(0.05)	(0.071)
Height	0.017	0.021	0.232	-0.09	-0.242	-0.435
	(0.816)	(0.333)	(0.552)	(0.337)	(0.335)	(0.445)
Weight	-0.28	-0.082	-0.083	-0.01	0.063	0.198
	(0.258)	(0.106)	(0.174)	(0.083)	(0.085)	(0.111)
Father's education	0.015	0.006	0.024	-0.016	-0.013	-0.025
	(0.044)	(0.019)	(0.03)	(0.01)	(0.011)	(0.013)
Mother's education	-0.036	-0.035	-0.065	-0.014	-0.039**	-0.05**
	(0.057)	(0.027)	(0.039)	(0.012)	(0.014)	(0.016)
Observations	1142	1142	1142	2647	2647	2647

Note: (1)(4) are the estimation by OLS. (2)(5) are the estimation by Poisson. (3)(6) are the estimation by ordered probit.
***p<0.001, **p<0.01, *p<0.05

Data from CGSS

We will discuss the results of each variable in all these specifications above.

4.2 Income

The first row of all tables shows the impact of income on the number of children. In the baseline model, the OLS results indicate that income has no significant impact on the number of children, and the corresponding Poisson regression results and ordered probit regression results also support this conclusion, that income has no significant impact on the

number of children.

In Table 2.3, only the sixth column showed significant results, which means that among non-rural holders, income has a significant positive effect on the number of children. However, the corresponding OLS and Poisson regression results are not significant, which makes the results in the sixth column not so convincing.

The impressive results are shown in Table 2.4. In columns 1, 2, and 3 of Table 2.4, there is a significant positive effect of income on the number of children. This series of results indicate that the income of respondents working in agriculture is indeed positively correlated with the number of children. This also partly confirms the hypothesis stated earlier that among those whose income growth depends not on the human capital accumulation such as agricultural work, their income is positively related to the number of children.

In summary, income generally has no significant impact on the number of children. If we investigate those who are only engaged in agricultural work, then we see that their income has a positive impact on their children. This shows that parenting strategies may be different in the people who are engaged in agricultural work and non-agricultural work.

We can indicate that one of the reasons why the two views in the introduction are contradictory is that the researchers used samples of different people in their empirical research. The different strategies of people in different samples lead to different correlations between income and the number of children.

4.3 Housing

There is a statistically significant positive effect of house size on the number of children in the results in all columns of Table 2.3. If we look at the results of Table 2.3, we find that the result in the first column is 0.169, less than the 0.263 in the fourth column. Similarly, the result in the second column is 0.077, less than the result in the fifth column,

which is equal to 0.122. The result in the third column is 0.136, less than the result in the sixth column which is equal to 0.299. These results indicate that in the samples that are categorized by hukou, house size has a statistically significant positive effect on the number of children. However, the impact of house size on the number of children for rural hukou holders is smaller than for non-rural hukou holders.

This pattern of differences between coefficients only exists in Table 2.3. We did not observe it in Table 2.4. As we can see in Table 2.4, the result of the first column is 0.245, which is larger than the 0.178 of the fourth column, the result of the second column is 0.09, less than the 0.096 of the fifth column, and the result of the third column is 0.145, less than the 0.213 of the sixth column. This means that for respondents engaged in agricultural work, and respondents engaged in non-agricultural work, the size of their houses both have a positive effect on the number of children, but we cannot consistently determine whose effect is larger.

One explanation for this is that housing policy in China is implemented according to hukou status, not job type. As we mentioned earlier, a rural hukou holder has the right to own property in rural areas, but not necessarily engage in local agricultural work. Similarly, non-rural hukou holders may engage in agricultural work or non-agricultural work, but they will not be able to legally own a property in rural areas. Therefore, when we split the sample based on differences in hukou status, we can observe differences in the coefficients of house size between the two samples. However, such differences are not pronounced when we split the samples based on differences in job.

In most of the tables, the number of houses does not have a statistically significant effect on the number of children, with only two exceptions. In the first column of Table 2.3, the number of houses has a significant negative effect on the number of children. In the sixth column of Table 2.3, the number of houses has a significant positive effect on the number of

children. Overall, we cannot definitively say that the number of houses has a statistically significant effect on the number of children.

4.4 Hukou

In Table 2.2, we can see that there is a statistically significant negative effect of hukou on the number of children. We can find similar results in Table 2.4. There is a statistically significant negative effect of hukou on the number of children in all the columns except for the first column. As hukou is designed as a dummy variable, where 0 represents rural hukou and 1 represents non-rural hukou, the negative regression coefficient implies that non-rural hukou holders have fewer children than rural hukou holders.

4.5 Job

In columns 1 and 3 of Table 2.2, we can see that the job has a statistically significant positive effect on the number of children. As the job is defined as a dummy variable, where 0 represents non-agricultural jobs and 1 represents agricultural jobs, positive coefficient means that respondents who engage in agricultural job have more children.

In columns 1, 2, and 3 of Table 2.3, jobs have no statistically significant effect on the number of children. In columns 4, 5, and 6, jobs have a statistically significant positive effect on the number of children. This means that in the sample of rural hukou holders, respondents' job has no statistically significant effect on the number of children, while in the sample of non-rural hukou holders, respondents who are engaged in non-agricultural jobs have fewer children.

In Tables 2.2 and 2.3, we can only observe the effect of the job on the dependent variable number of children. The effect of the job on other variables is mainly reflected in Table 2.4, and we will discuss those effects as we discuss those variables.

4.6 Education spending

We did not find a statistically significant impact of education spending on the number of children in Table 2.2. In columns 4 and 6 of Table 2.3, and column 4 of Table 2.4, education spending has a positive statistically significant impact on the number of children. This empirical result contradicts the mainstream microeconomic theory. As we know from microeconomic theory, an increasing price leads to a decreasing demand, thus an increase in education spending raises the cost of raising children, which should lead to a reduction in the number of children. There are several possible explanations for this empirical result.

One possible explanation is that families may choose a strategy of increasing the number of children while also increasing the educational investment for each child. Such an idea may reflect an "investment" concept, where children are not only seen as part of the family, but also seen as an investment that could bring long-term returns. Parents may not only be willing to invest more resources in education and child-rearing, but also be willing to have more children, and provide each child with higher education resources. Such a strategy would reduce parents' welfare in the short term, but if parents expect that this strategy can bring them higher returns in the long term, then this strategy can be economically rational.

Another possible explanation is that wealthier families tend to buy more childcare services from the market, thereby improving the efficiency of childcare. Rich families may have more resources to buy childcare services, such as hiring a nanny, a tutor, or buying other services. In this way, they can effectively manage and allocate their time and resources, and they can afford the cost of more children, even when increasing education spending. This might be a possible explanation for the positive correlation between the increase in education spending and the number of children.

In future research, we can introduce "spending on education per child" mentioned in

data description of this chapter as a new variable to improve the analysis results. Incorporating “spending on education per child” into the analysis can help us explore whether the relationship between education spending and the number of children varies with the different investment amounts in education per child. This approach can reveal differences in how families allocate educational resources and how these differences affect their fertility behavior.

For example, if a family invests more in education per child, it may indicate a preference for having fewer children but providing better educational resources for each child. Conversely, if a family invests less in education per child, it might reflect a preference for having more children with less educational resources available to each child.

Moreover, spending on education per child can also serve as an indicator of a family’s economic status and priority for educational spending. Higher spending on education per child may be positively correlated with a family’s economic capability and the value they place on education. By analyzing the relationship between spending on education per child and the number of children, we can gain deeper insights into the interplay between family economic status, educational investment decisions, and fertility behavior. Such analysis can reveal how families at different economic levels respond to rising education costs and how these responses affect their fertility choices.

In practical research, incorporating spending on education per child as an analytical variable may require collecting more comprehensive data, including total family education spending, family size, and family structure. This data can help us more accurately calculate spending on education per child and analyze its relationship with the number of children. Additionally, considering that education spending may vary significantly with the child’s age and educational stage, the research should also consider the impact of these factors on spending on education per child. For example, the costs of early childhood education are

usually higher than those of the basic education stage, which may affect families' decisions on educational investments for children of different ages, thereby influencing fertility behavior (Heckman, 2011).

4.7 Age

Almost all results show that there is a statistically significant positive correlation between age and the number of children, which means that older respondents have more children. We can simply explain that the result is due to the elderly people having lived for a longer time than young people. However, there is another possibility that young people have less children because young people really have lower fertility propensity than last generation. This possibility does exist, but we cannot directly prove it from this empirical result.

4.8 Education

In most of the results across all tables, there is a statistically significant negative correlation between respondents' education level and the number of children, except for columns 4 and 5 of Table 2.3 and column 3 of Table 2.4. This is consistent with the classic findings of previous demographic studies, that is, there is a negative relationship between family size and the increase in respondents' education level, which means the higher the respondents' education level, the fewer children they tend to have. The effect of education on fertility is explored in detail in the preceding chapter.

4.9 Gender Difference

In all the three results of Table 2.2, there is a positive correlation between gender and the number of children. Since we define gender as a dummy variable, where male is defined as 0, and females are defined as 1, the positive regression coefficient means that in the sample,

women reported to have more children than men. Biologically speaking, in a sufficiently large sample, males and females should have the same number of children, but our empirical results do not show this.

One explanation is that the children in this survey are not necessarily biological children, but rather children in a social or legal sense. Some children born out of wedlock might be raised more by females, and children might also be handed over to females after a couple is divorced, so these children's biological fathers do not regard these children as their children in the social or legal sense.

4.10 Ethnicity and Religion

In most of the results, there is no statistically significant effect of ethnicity difference on the number of children, except for column 6 of Table 2.3. As we see in Table 2.1, the variable ethnicity has a total of 8 discrete values, where 1 represents the Han, and 2 to 7 represent the other ethnic groups. The mean value of ethnicity is 1.348, implying that the Han is the majority ethnic group in the sample. Therefore, the positive regression coefficient in column 6 of Table 2.3 represents that the Han people have fewer children compared to other ethnic groups.

In all the three results of Table 2.2, there is a statistically significant negative correlation between religion and the number of children. We also find statistically significant negative correlations in columns 1, 2, 3, and 4 of Table 2.3, and columns 3, 4, 5, and 6 of Table 2.4. Since we define religion as a dummy variable, where having a religious belief is defined as 0, and having no religious belief is defined as 1, the negative statistical coefficient means that people with religious beliefs have more children than people with no religious belief. This result is consistent with previous empirical research. Our empirical results also indicate that statistical significance appears more in the sample of respondents having rural

hukou and the sample of respondents having non-agricultural jobs.

4.11 Height and Weight

In all the results across all tables, there is no statistically significant relationship between the respondents' height and weight and the dependent variable.

4.12 Father's and Mother's Education

In most of the results, there is no statistically significant relationship between the number of respondents' children and the education level of respondents' father, except for column 6 of Table 2.3, where the education level of the respondents' father has a negative statistically significant impact on the number of respondents' children.

Compared to the education level of respondents' father, the education level of the respondents' mother shows more significance in the results. In columns 2 and 3 of Table 2.2, columns 2, 3, 4, 5, and 6 of Table 2.3, columns 5 and 6 of Table 2.4, the education level of the respondents' mother has a statistically significant negative impact on the respondents' children.

As we previously mentioned, considering the tradition in China of grandparents raising children, we have added the education level of respondents' parents as control variables. If we assume that grandparents actually participate in their grandchildren raising, our empirical results are consistent with the previous demographic studies, which means that the decrease in the number of children is influenced by the improvement in female education level. Meanwhile the impact from the improvement of respondents' parents' education level is less than the impact of the respondents' own education level on the number of respondents' children (Cai, 2023).

4.13 Policy

To analyze the impact of policy, we select three representative years, 1971, 1982, and 1991. As we mentioned in the earlier section, the Chinese government issued or strengthened the one-child policy in these years. We created dummy variables named Policy 1, Policy 2, and Policy 3. For each variable, if the respondent's birth year is earlier than or equal to the year the policy was issued, the corresponding value of the dummy variable is set to 0. On the contrary, if the respondent's birth year is later than the year the policy was issued, the corresponding value of the dummy variable is set to 1. Based on the definition of dummy variables, these variables do not represent the direct impact of the policy on the fertility rate in the year it was issued but represent the impact of the policy on the fertility rate of the new generation born after the policy was issued. Therefore, what we are studying here is the indirect impact of the policy.

In Table 2.5, we can see that Policy 1 has a significant positive effect on the dependent variable in all three columns, while Policy 2 only has a significant positive effect on the dependent variable in the first column. Like Policy 1, Policy 3 has a significant impact on the dependent variable in all three columns, but its impact is negative. This empirical result implies that the effect of Policy 1 on the fertility rate of the generation born after the year of Policy 1 issued is positive, and the effect of Policy 3 on the fertility rate of the generation born after the year of Policy 3 issued is negative.

This is an interesting result, because the purpose of all three policies is to reduce fertility rates. Therefore, we expect a negative correlation between the policy and the dependent variable, but we found a positive correlation in our empirical findings. One possible explanation is that policies have a lag effect. People did not fully accept the policy when it was first issued, and therefore the fertility rate remains on a high level. This could be a possible reason for the positive correlation between Policy 1 and the dependent variable.

Table 2.5 Results of Policy Effect

	Dependent variable: number of children		
	(1)	(2)	(3)
Income	0.002	0.001	0.004
	(0.008)	(0.005)	(0.008)
House size	0.202***	0.09***	0.189***
	(0.019)	(0.01)	(0.019)
Houses	-0.003	0.002	0.037
	(0.027)	(0.019)	(0.027)
Hukou	-0.4***	-0.252***	-0.543***
	(0.048)	(0.035)	(0.05)
Job	0.251***	0.07*	0.18***
	(0.05)	(0.033)	(0.049)
Spending on education	0.004	0.001	0.004
	(0.004)	(0.002)	(0.004)
Age	0.052***	0.024***	0.051***
	(0.002)	(0.001)	(0.002)
Education	-0.031***	-0.024***	-0.044***
	(0.008)	(0.007)	(0.009)
Gender	0.125*	0.071*	0.176***
	(0.049)	(0.033)	(0.049)
Ethnicity	0.009	0.005	0.02
	(0.013)	(0.009)	(0.013)
Religion	-0.219***	-0.108**	-0.225***
	(0.059)	(0.038)	(0.059)
Height	0.042	-0.046	0.012
	(0.347)	(0.236)	(0.349)
Weight	-0.113	-0.032	0.017
	(0.093)	(0.066)	(0.094)
Father's education	-0.012	-0.011	-0.02
	(0.012)	(0.01)	(0.012)
Mother's education	-0.015	-0.04**	-0.064***
	(0.014)	(0.013)	(0.015)
Policy 1	0.66***	0.301***	0.642***
	(0.067)	(0.048)	(0.068)
Policy 2	0.148*	-0.053	-0.075
	(0.071)	(0.059)	(0.073)
Policy 3	-0.324***	-1.388***	-1.401***
	(0.084)	(0.142)	(0.11)
Observations	3789	3789	3789

Note: (1) is the estimation by OLS. (2) is the estimation by Poisson. (3) is the estimation by ordered probit. ***p<0.001, **p<0.01, *p<0.05

Data from CGSS

After a period, when people gradually accept the policy, the impact of the policy begins to show up and fertility rates start to decrease. Therefore, there is a negative correlation between

Policy 3 and the dependent variable. The fact that Policy 2 does not show a significant effect could be an indirect evidence of policy lag.

5. Conclusion

The positive correlation between income and fertility rate indeed exists, but it is only in the sample of the people who are engaged in agriculture jobs. There is no statistical correlation between income and fertility rate for those engaged in other jobs.

Compared to income, there is a very clear positive correlation between house size and the number of children, while there is no correlation between the number of houses and the number of children.

Hukou and jobs have multifaceted effects on the number of children. When Hukou and jobs are considered as index variables to seperate the sample, rural Hukou holders and agricultural workers have more children.

There is basically no statistical correlation between the number of children and the net cost of children as measured by education spending. Only a few results show a significant positive correlation. Such empirical results indicate that special child-rearing strategies may exist in some families that are more under traditional Chinese cultural influence.

There is a negative correlation between the education level of respondents and the number of their children, which is a result consistent with previous research findings and the preceding chapter. Additionally, there is a statistically significant correlation between the education level of respondents' parents and the number of respondents' children.

Different groups of people may have very different fertility strategies, so it is necessary to identify different samples. The final fertility strategy of the population might be a mixture of various strategies.

References

Becker, Gary S. 1960. “An Economic Analysis of Fertility.” *Universities-National Bureau*, p. 209 - 240.

Borg, Mary O'Malley. 1989. “The Income-Fertility Relationship: Effect of the Net Price of a Child.” *Demography*, Vol. 26, No.2.

Cameron, L., N. Erkal, L. Gangadharan, and X. Meng. 2013. “Little Emperors: behavioral Impacts of China's One-Child Policy.” *Science*, VOL. 339, NO. 6122.

Chen, Chuanbo and C. Cindy Fan. 2016. “China’s Hukou Puzzle: Why Don’t Rural Migrants Want Urban Hukou?” *China Review*, Volume 16, Number 3, October 2016 pp. 9-39.

Cheng, Yuan, and Lei Zhang. 2015. “Public education spending and private substitution in urban China.” *Journal of Development Economics*, 115 (2015) 124–139.

Cai, Fang. 2011. “Hukou System Reform and Unification of Rural–urban Social Welfare.” *China & World Economy*, Volume19, Issue3.

Cai, Lichun. 2023. “The Effect of Education Differences on Fertility Rate.”

Dyson, Tim. 2001. “World Food Trends: A Neo-Malthusian Prospect?” *Proceedings of the American Philosophical Society*, Dec. 2001, Vol. 145, No. 4 (Dec. 2001), pp. 438-455.

Ehrlich, Isaac and Jinyoung Kim. 2005. “Endogenous fertility, mortality and economic growth: Can a Malthusian framework account for the conflicting historical trends in population?” *Journal of Asian Economics*. Volume 16, Issue 5, October 2005, Pages 789-806.

Ewer, Phyllis A. and Eileen Crimmins-Gardner. 1978. “Income in the Income and Fertility Relationship.” *Journal of Marriage and Family*, May 1978, Vol. 40, No. 2 (May 1978), pp. 291-299.

Fu, Qiang and Qiang Ren. 2010. “Educational Inequality under China’s Rural–Urban Divide: The Hukou System and Return to Education.” *Environment and Planning A: Economy and Space*, Volume 42, Issue 3.

Galor, Oded, and David N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *The American Economic Review*, Vol. 90, No. 4 (Sep. 2000), pp. 806-828.

Heckman, James J. 2011. "The Economics of Inequality: The Value of Early Childhood Education." *American Educator*, v35 n1 p31-35, 47 Spr 2011.

Jones, Larry E., Alice Schoonbroodt and Michèle Tertilt. 2008. "Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?" *NBER Working Paper*, No. 14266.

Kalemli-Ozcan, Sebnem. 2003. "A stochastic model of mortality, fertility, and human capital investment." *Journal of Development Economics*, Volume 70, Issue 1, February 2003, Pages 103-118.

Kögel, Tomas and Alexia Prskawetz. 2001. "Agricultural productivity growth and escape from the Malthusian trap." *Journal of economic growth*, 6, 337-357, 2001.

Kulu, Hill and Andres Vikat. 2007. "Fertility differences by housing type." *Demographic Research*, JULY - DECEMBER 2007, Vol. 17 (JULY - DECEMBER 2007), pp. 775-802.

Kulu, Hill. 2010. "Why Do Fertility Levels Vary between Urban and Rural Areas?" *Regional Studies*,Volume 47, 2013 - Issue 6: Financialization, Space and Place.

Kunz, Phillip R. 1965. "The Relation of Income and Fertility." *Journal of Marriage and Family*, Nov. 1965, Vol. 27, No. 4 (Nov. 1965), pp. 509-513.

Lerch, Mathias. 2019. "Fertility Decline in Urban and Rural Areas of Developing Countries." *Population and Development Review*, Vol. 45, No. 2 (JUNE 2019), pp. 301-320.

Liu, Zhiqiang. 2005. "Institution and inequality: the hukou system in China." *Journal of Comparative Economics*, Volume 33, Issue 1, March 2005, Pages 133-157.

Martin, Teresa Castro. 1995. "Women's Education and Fertility: Results from 26 Demographic and Health Surveys." *Studies in Family Planning*, Jul. - Aug. 1995, Vol. 26, No. 4 (Jul. - Aug. 1995), pp. 187-202.

Ohno, Susumu. 1996. "The Malthusian parameter of ascents: What prevents the exponential increase of one's ancestors?" *Proc. Natl. Acad. Sci. USA* 93 (1996).

Rosenzweig, Mark R. and Kenneth I. Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica*, Vol. 48, No. 1 (Jan. 1980), pp. 227-240.

Shen, Ling. 2012. "Are house prices too high in China?" *China Economic Review*,Volume 23, Issue 4, December 2012, Pages 1206-1210.

Schultz, T. Paul. 1990. "Testing the Neoclassical Model of Family Labor Supply and Fertility." *The Journal of Human Resources*, Autumn, 1990, Vol. 25, No. 4 (Autumn, 1990), pp. 599-634.

Schultz, T. Paul. 2005. "Fertility and Income." *Center Discussion Paper*, NO. 925.

Tournemaine, Frederic and Pongsak Luangaram. 2010. "R&D, human capital, fertility, and growth." *J Popul Econ* (2012) 25:923–953 DOI 10.1007/s00148-010-0346-4.

Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and Its Effects on Family Outcomes." *Journal of Economic Perspectives*, Volume 31, Number 1—Winter 2017—Pages 141–160.

Zhu, Huiming, Zheng Li and Peng Guo. 2018. "The impact of income, economic openness and interest rates on housing prices in China: evidence from dynamic panel quantile regression." *Applied Economics*,Volume 50, 2018 - Issue 38.

III. The Impact of Activities on Low Fertility Rate: Evidence from China

1. Introduction

As the most populous country in the world, China's fertility rate has decreased dramatically in recent years. China's fertility rate peaked in 1987 and then has continued to decline. In 2016, China government abolished the one-child policy that last for nearly 40 years and implemented a comprehensive two-child policy (Wang, Gu and Cai, 2016). However, what was unexpected was that the policy did not achieve the expected results. In 2017 there were only 17.23 million births, with a decrease of 630,000 over 2016. The fertility rate continued to decline again after 2017. In 2020, the desired number of children of childbearing age women was 1.8, while the actual total fertility rate was only 1.3, and the number of newborns registered by the public security department was only 10.035 million, which was a sharp decrease of 15% from 11.79 million in 2019 (Yu, 2021).

The results of the seventh national population census show that China's population continues to grow, but it is difficult to say whether this growth comes from the increase in life expectancy or the increase in the number of newborns. The decline in the number of newborns in China will inevitably have a negative impact on China's labor supply in decades to come, thus weakening China's economic vitality. Although the dramatic decline of China's fertility rate cannot be said to have become a public crisis, it can be surely predicted that it will bring serious challenges to China's long-term economic growth in the future, and may be one of the most core challenges (Dong, 2021). The motivation of this paper is to explore the relationship between the number of children and the impact of individual social activities, to provide valuable suggestions for public policy makers.

The factors affecting population growth have always been one of the key issues in population economics studies. The discussion on population growth can be summarized into two theoretical frameworks: Malthusian theoretical framework and non-Malthusian

theoretical framework. The key point of Malthus's theory is that the population always has the trend of exponential growth, and the output can only grow linearly, so the total population is limited by the growth of output. Malthus tried to describe a world in which humankind tried their best to reproduce, but the lack of output in that world always put humankind in a tragic situation (Galor and Weil, 2000).

In the history of the pre-industrial era, Malthus's theory, if not completely correct, is considered to have a wide influence, so that it has a significant impact on the population policies of countries in the industrial era. Similarly, Malthus's theory lacks explanatory power in the era after countries are industrialized, particularly after the end of World War II. After World War II, a baby boom took place all over the world, and then the total fertility rate went all the way down, which provided a background for the birth of various non-Malthusian theories. Non-Malthusian theory no longer presupposes human reproductive habits but turns to seek empirical connections from a variety of factors such as the decisive role of education (Kunz, 1965).

The decisive effect of the improvement of education level on the low fertility rate has become a classic economic explanation, that is, the improvement of education level, particularly women's education level significantly reduces the fertility rate. The internal logic of this view is that the direct result of the improvement of education level increases the childcare cost. Spending more time on education inevitably delays the age of marriage and reduces the time available for childcare, resulting in the decline of fertility (Martin, 1995).

Malthusians have some criticism on educational determinism. They believe that the improvement of women's education level is accompanied by the increase of output of the whole economy, which makes individuals, families and societies more afford to have more children, unless people do not want children subjectively (Mellos, 1988). If Malthusians are correct, considering the sustained economic development in the post-war world, the

population explosion will continue after the post-war baby boom, but in fact, there is no such population explosion, and even the population does not grow linearly with the economic growth. This has at least proved that the traditional Malthusian theory cannot fully explain the demographic phenomenon of industrialized society.

The criticism of educational determinism does not lie in its defects in empirical analysis. The negative correlation between the improvement of women's education level and the decline of fertility is statistically convincing, but it does not provide us with sufficient theoretical details. Similarly, there are other explanations, such as the impact from income, religion, urbanization and other factors. All theories that emphasize the role of a single factor in low fertility have the same problem, that is, they are statistically convincing but difficult to understand in theoretical details. We want to know in all economies with improved education, whether there are more specific and observable changes which are related to the improvement of education and have a more direct impact on fertility rate or fertility choices?

These changes have occurred at least in parts of East Asia. This is a change in marriage and fertility culture in East Asia. One of the most famous examples is the "Otaku" generation in Japan. Otaku refers to a group of people who are addicted to various subcultures and have an extremely in-depth understanding of the subculture (Kinsella, 1998). In a narrow sense, it refers to people who are addicted to animation, comics and video games. The obvious feature of this group is their lack of social interaction, which keeps them from falling into love, marriage and childbirth. Similarly in South Korea, there has been a group of people called "Sampo generation", which means to abandon the tradition of falling into love, marriage and childbirth (Muradyan, Yashkina, 2020).

Like her neighbors in East Asia such as Japan, South Korea and Singapore, China's economic growth has entered a relatively stable period after experiencing a period of rapid economic growth in the early 21st century. At present, it is uncertain whether there are

similar social groups in China as in other parts of East Asia, but similar cultural phenomena have emerged. The popular concept of "TangPing"[5] on the Chinese Internet is one of them which means to earn only the minimum wage, not to invest more in human capital, and not to engage in social activities including falling in love, marriage and childbirth.

Phenomena such as "Otaku", "Sampo generations" and "TangPing" can be summarized into a common feature, that is, the relevant people show a lack of social activities and interaction with other people. If the activity of these people contains factors that can explain the low fertility rate in East Asia, this factor is more likely to belong to a non-Malthusian framework.

However, how to explain a cultural phenomenon, such as the lack of correlation between social activities and fertility? Having children can bring benefits to parents. The composition of utility of having children is complex, including but not limited to economic utility, social status utility and psychological satisfaction utility.

If an individual can obtain goods that provide physical and mental satisfaction at a lower cost, the desire to have children will be reduced, and in extreme cases, there may be even no need for children at all. In other words, children and some other goods are substitutes for each other. The new lifestyle created by modern society provides many similar options, such as mobile phones and the Internet. These goods also have a common feature, that is they do not need to rely on social interaction in physical sense.

Individual fertility depends on the fertility activity of other members of the population, and on the structure of the interactions between individuals. If we assume that individual beliefs and activities are interdependent and are moderated by social interaction mechanisms and social structures, then social mechanisms, such as social learning, social pressure, social contagion, and the social exchange of resources would affect individuals' beliefs and norms

[5] It means "lying flat on the ground" in Chinese.

regarding childbearing, and the actual and perceived opportunities and constraints which shape their childbearing choices (Bernardi and Klärner, 2014).

A key question is, how to measure an activity to determine whether it is social or non-social? Basically, it is completely subjective to determine whether an activity belongs to social activities or non-social activities. We can subjectively compare the difference in social extent between two activities, and we should not only consider in narrow sense "social activities" such as making friends and falling in love, but also consider the alternatives of social activities, especially the emerging lifestyles and "non-social" activities in the industrial era. These non-social activities include but are not limited to being alone, reading, watching TV or movies, playing video games, single player sports and so forth. Given the limited time, spending more time on non-social activity means spending less time on social activity thus no benefits to the formation of marriage and childbirth activity. Marriage and childbirth activity are the result of social activity, not just the output of human biological function or simple economic interests. People must first have contact with others and carry out some social activities before they can fall in love, even enter marriage and childbirth. Furthermore, being in a non-social environment for a long time may make people lose their social interaction ability.

Among all the non-social activities, the activity on the Internet is the most complex because it is difficult to determine whether Internet activities belong to social activities or non-social activities. Sometimes people's objective of surfing the Internet is to contact or socialize with others. This kind of socializing is not face-to-face, but there is still a lot of effective information exchange. Some activities on the Internet are obviously social, while others are obviously not, or although these activities are related to the Internet, there are obvious differences in social extent.

For example, the activity of surfing the Internet such as browsing the web alone is

obviously a non-social activity, while chatting with other people on social media is an obvious social activity. If we can get more similar details in the data, it will undoubtedly help to clarify many concepts. But overall, it is difficult to define the relevant activity, especially in the Internet community. Our criterion of social activity is clear, that is the activity that needs human face-to-face contact and not in virtual mode. Perhaps people can really meet and fall in love through the Internet, but in terms of current technology, it is impossible for people to have children through the Internet. Similarly, playing video games is also classified as a non-social activity. Although many types of video games need multi-players interaction, this interaction is undoubtedly weaker than the traditional large-scale multi person outdoor activities.

Another question is how to define the activity done in public, such as watching a movie in a cinema or watching a sports game on site? Is the activity done in public equal to social activity? Our criterion is whether an activity is judged as a social activity, not only according to whether it is done physically in a public area, but more importantly, whether there is information exchange from others on the scene. In terms of this criterion, watching a movie in a cinema alone is basically a non-social activity, while watching a sports game on site is a social activity. Because, when watching movies in cinemas, although all audiences are in public places and there is a large amount of information exchange among all the audience and the screen, this exchange is generally one-way from the screen to the audience. And most importantly, there is generally no information and interaction exchange among all the audiences. Therefore, watching movies in the cinema is a typical non-social activity in public areas. Watching a sports game on site is another case. When watching the game, even if the audiences are not sending effective information, they are engaged in social activities such as cheers and sharing feelings with others. Therefore, watching a sports game on site is regarded a social activity.

The hypothesis of this paper is that social activity has a positive impact on fertility and non-social activity has a negative impact on fertility. There are two dependent variables in the model, the ideal number of children and the current number of children. The policy restrictions here mainly refer to China's "one child" policy which was abolished in 2016. The "one child" policy was replaced by the "two children" policy, that is, a couple can have two children. In 2021, the "two children" policy was abolished again and replaced by the "three children" policy, that is, a couple can have up to three children. In terms of the details of the new policy, the "three children" policy is much looser than the "two children" policy. China's fertility policy has changed from restricting fertility to encouraging fertility. Considering that the "one child" policy has been implemented for 34 years since 1982, which has profoundly changed China's fertility culture, the impact of these policies cannot be ignored in demographic research.

The independent variables in the model will be identified into two parts: "basic information" or traditional variables such as education and income, and the "individual activity" such as various social activities and non-social activities. Variables such as province and urbanization are treated as control variables. The model will use two samples, one is the sample containing all observations, and the other is the sample filtered by age, retaining the observations born after 1981. The main contribution is to verify the factors affecting fertility intention in previous studies, and to explore the impact of social activities and non-social activities on fertility intention.

2. Data and model-relevant facts

The data used in this paper is China General Social Survey (CGSS). China comprehensive social survey is a survey jointly conducted by Renmin University of China and other academic institutions all over China in every odd year since 2003. This survey

covers all aspects of life including work, family and so on. This survey gives us a random sample of cross-sectional data, including 783 variables and 12582 observations. This study used the last data from 2017.

CGSS consists of four sections A, C, D and Z. Among them, section A is the core module, including 90 questions, covering all the basic personal information. section C contains 66 questions mainly about social activities. section D contains 42 questions mainly to further describe the social relationship of the respondents. section Z contains 9 questions, which is a supplement to the personal information of the respondents.

We selected some answers to the questions in the original survey to form a data set for the study. There are some missing values in the dataset. All the observations associated with the missing values are dropped. After clearing all observations with missing values, we obtained a sample with 11595 observations. Our next data processing work will be carried out on this new sample.

Then for some observations, the value of "I don't know" is set to 98 or 998, and the value of "I refuse to answer the question" is set to 99 or 999. These kinds of observations are dropped. After the raw data is cleaned up, Table 3.1 shows the mean, standard deviations, maximum value and minimum value of all variables.

The dependent variable is "ideal number of children" and "current number of children". The value of the ideal number of children comes from the survey question A37 "how many children do you want if there are no policy restrictions?" This question corresponds to three answers "how many children do you want to have", "how many sons do you want to have" and "how many daughters do you want to have". The first answer to this question is chosen to be the value of the ideal number of children.

The variable "current number of children" comes from the survey question A68 "How many children do you have, including stepchildren, adopted children, and children passed

Table 3.1 Descriptive Statistics

	Mean	SD	Min	Max
Ideal number of children	1.963	1.073	0	21
Current number of children	1.701	1.360	0	32
Province	-	-	1	31
Gender	0.4721	0.499	0	1
Ethnicity	-	-	1	8
Religion	1.412	1.243	1	9
Education	5.133	3.275	1	14
Income	36827	196743.8	0	9993600
House size	110	102.128	0	2400
Height	162.8	15.172	0	193
Weight	120	29.370	0	260
Health	3.452	1.110	0	5
Urbanization	-	-	1	7
Working hours per week	1.165	8.080	0	140
Unemployment months	40.9	87.014	0	936
Number of family members	2.806	1.536	0	44
Reading newspaper	1.761	1.117	0	5
Reading magazine	1.596	0.899	0	5
Listening to broadcast	1.741	1.108	0	5
Watching TV	3.784	1.128	0	5
Sending text	2.787	1.721	0	5
Relaxing	3.434	0.895	0	5
Studying	1.955	1.109	0	5
Social with neighbors	4.083	2.207	0	7
Social with friends	4.02	1.873	0	7
Watching movie in cinema	4.494	0.773	0	5
Shopping	3.444	1.118	0	5
Attending concert	4.571	0.782	0	5
Party with family members	3.844	0.752	0	5
Party with friends	3.628	1.019	0	5
Listening to music at home	3.432	1.544	0	5
Sport exercise	3.515	1.596	0	5
Watching sports games on site	4.783	0.572	0	5
Doing manual artwork	4.571	0.899	0	5
Surfing Internet	2.972	1.891	0	5
Books read in last 12 months	4.951	22.612	0	990
E-books read in last 12 months	2.408	18.422	0	990
Newspaper and magazine read in last 12 months	35.83	110.016	0	996
Read online information in last 12 months	62.27	102.845	0	990
Having a cell phone	0.8908	0.312	0	1
Being online in last 6 months	0.5627	0.496	0	1
Other family member being online in last 6 months	0.7545	0.430	0	1
Wechat Pay in last 6 months	0.3962	0.489	0	1
Alipay in last 6 months	0.3242	0.468	0	1
		$N = 11595$		

Data from CGSS

away?" The question asks respondents to provide two answers, the number of sons and the

number of daughters. The actual variable used in regression is the sum of the two answers.

To make the regression results easy to interpret, the two dependent variables are magnified by 100 times, which will not affect the regression results.

The independent variables can be identified into two sections, the basic information and the individual activity. The basic information includes the following variables: province, gender, ethnicity, frequency of religious activity, current highest education level, total annual income in the last year, house size, height, weight, subjective health status, urbanization, weekly working hours, maximum unemployment time and number of family members. Province, ethnicity and urbanization are the control variables.[6]

Gender is a dummy variable, where the value of female is set to 0 and the value of male is set to 1. Some studies suggest that men and women play different roles in fertility decisions, with women potentially playing a crucial role. Therefore, discussing the original sample by separating men and women seems necessary. However, this view is not always accurate. In some cases, men might have an influence on fertility decisions, and even more so than women. Derose and Ezeh (2007) found that while lower fertility is commonly associated with women's reproductive autonomy, the influence of men's education on reproductive decision-making increased during the first decade of rapid fertility decline.

Other evidence indicates that fertility decisions may be a result of interactions between men and women, rather than a decision made by either gender. The decision to have a first child is made jointly, and each partner may exercise a veto (Bauer and Kneip, 2013). We mentioned that the mean number of children in the data is 1.701, which is a relatively low fertility level. Therefore, we can assume that the fertility decisions discussed in this data primarily concern the decision to have the first child. Hence, considering men's decisions in this data is reasonable.

[6] Attentive readers may notice the absence of a discussion about the marriage status in the control variables. This is simply because the survey did not contain relevant information.

Men's influence in fertility decisions might be related to their income. Kaufman and Bernhardt (2012) indicate that men are more likely to intend to have a child if their partner's job makes it easy to take parental leave or work part-time. Women are more likely to intend to have a child if their partner's job pays well. In addition, men whose job pays well are more likely to have a child.

Technically, the original survey is a random cross-sectional sample, so if we assume that there are no common children between any two respondents in the data, we can then approximate that each respondent represents a household. This means that the decision of the respondent represents the household's decision.

Individual activity variables are the following: the frequency of reading newspapers last year, the frequency of reading magazines last year, the frequency of listening to broadcast last year, the frequency of watching TV last year, the frequency of sending text last year, the frequency of relax time in last year, and the frequency of studying last year. The range of these six variables is 1 to 5, which means the frequency from low to high, where the value 1 means "never" and value 5 means "very frequent". The range of the two variables "social with neighbors" and "social with friends" is 1 to 7, where value 1 means "almost every day" and value 7 means "never". For the variables of "watching movies in the cinema", "go shopping", "attending concerts", "party with family members", "party with friends", "listen to music at home", "sport exercise", "watching sports games on site", "doing manual artwork" and "surfing Internet", the value range is 1 to 5, representing the frequency from high to low, where the value 1 means "everyday" and value 5 means "never". The variables "books read last year", "e-books read last year", "newspapers and magazines read last year", and "read information online last year" are numerical. In the last section there are five dummy variables, "having a mobile phone", "being online in the last 6 months", "other family members being online in the last 6 months", "used Wechat Pay in the last 6 months" and "used Alipay in the

last 6 months".

Among all the independent variables in the section of individual activities, we consider these activities as social activities: "social with neighbors", "social with friends", "shopping", "party with family members", "party with friends" and "watching sports games on site". And the rest variables of individual activities are non-social activities. Our hypothesis is that these social activity variables have a positive impact on the two dependent variables, while non-social activities have a negative impact on the two dependent variables.

3. Models

There are 4 different models built based on the variables mentioned in the last section, they are baseline models by applying OLS regression, Ridge model and Lasso model.

The baseline models are:

$$inc_i=\beta_0+\beta_1 pro_i+\beta_2 sex_i+\beta_3 nat_i+\beta_4 rel_i+\beta_5 edu_i+\beta_6 icm_i+\beta_7 hsz_i+\beta_8 hei_i+\beta_9 wei_i+\beta_{10} hel_i+\beta_{11} urb_i+\beta_{12} whw_i+\beta_{13} umt_i+\beta_{14} nfm_i+\beta_{15} rnp_i+\beta_{16} rmz_i+\beta_{17} lbc_i+\beta_{18} wtv_i+\beta_{19} stx_i+\beta_{20} rlx_i+\beta_{21} std_i+\beta_{22} swn_i+\beta_{23} swf_i+\beta_{24} wmc_i+\beta_{25} shp_i+\beta_{26} atc_i+\beta_{27} pfm_i+\beta_{28} pwf_i+\beta_{29} lmh_i+\beta_{30} spx_i+\beta_{31} wss_i+\beta_{32} dmw_i+\beta_{33} sln_i+\beta_{34} rbl_i+\beta_{35} ebl_i+\beta_{36} nml_i+\beta_{37} roi_i+\beta_{38} hcp_i+\beta_{39} bol_i+\beta_{40} fmo_i+\beta_{41} wcp_i+\beta_{42} alp_i+\varepsilon_i \quad (3.1)$$

$$cnc_i=\beta_0+\beta_1 pro_i+\beta_2 sex_i+\beta_3 nat_i+\beta_4 rel_i+\beta_5 edu_i+\beta_6 icm_i+\beta_7 hsz_i+\beta_8 hei_i+\beta_9 wei_i+\beta_{10} hel_i+\beta_{11} urb_i+\beta_{12} whw_i+\beta_{13} umt_i+\beta_{14} nfm_i+\beta_{15} rnp_i+\beta_{16} rmz_i+\beta_{17} lbc_i+\beta_{18} wtv_i+\beta_{19} stx_i+\beta_{20} rlx_i+\beta_{21} std_i+\beta_{22} swn_i+\beta_{23} swf_i+\beta_{24} wmc_i+\beta_{25} shp_i+\beta_{26} atc_i+\beta_{27} pfm_i+\beta_{28} pwf_i+\beta_{29} lmh_i+\beta_{30} spx_i+\beta_{31} wss_i+\beta_{32} dmw_i+\beta_{33} sln_i+\beta_{34} rbl_i+\beta_{35} ebl_i+\beta_{36} nml_i+\beta_{37} roi_i+\beta_{38} hcp_i+\beta_{39} bol_i+\beta_{40} fmo_i+\beta_{41} wcp_i+\beta_{42} alp_i+\varepsilon_i \quad (3.2)$$

Each regression equation estimates two datasets, so there are four estimation results in the end as shown in Table 3.2. In the first model, the dependent variable is the ideal number of children, and the dataset includes all observations. In the second model, the dependent

Table 3.2 Results of Baseline Model

	Dependent variable:			
	inc		cnc	
	(1)	(2)	(3)	(4)
Province	0.6766***	0.5201*	0.649***	0.07075
	(0.1089)	(0.2032)	(0.1202)	(0.1587)
Gender	3.085	6.081	3.297	-19.67***
	(2.266)	(4.593)	(2.501)	(35.87)
Ethnicity	1.015	1.286	-0.1156	-2.128*
	(0.7033)	(1.318)	(0.7764)	(1.03)
Religion	5.092***	6.771***	4.981***	1.364
	(0.7917)	(1.745)	(0.8739)	(1.363)
Education	-1.06*	0.1579	-5.207***	-3.506***
	(0.452)	(0.7044)	(0.4989)	(0.5502)
Income	9.376E-07	7.208E-05**	6.01E-06	1.337E-04***
	(4.969E-06)	(2.195E-05)	(5.49E-06)	(1.714E-05)
House size	0.04218***	0.02814	0.07273***	0.01929
	(0.009887)	(0.01852)	(0.01091)	(0.01447)
Height	-0.1403	-0.2759	-0.3672***	0.02429
	(0.07397)	(0.177)	(0.08166)	(0.1382)
Weight	0.009669	0.1065	-0.0566	0.2292***
	(0.03774)	(0.06714)	(0.04166)	(0.05244)
Health	0.6896	-2.044	-7.453***	-1.499
	(0.9885)	(2.191)	(1.091)	(1.712)
Urbanization	-6.759***	-4.483*	-9.069***	-0.8489
	(1.029)	(1.864)	(1.136)	(1.456)
Working hours per week	0.06872***	0.158	-0.04092	0.1157
	(0.1209)	(0.1705)	(0.1335)	(0.1331)
Unemployment months	0.04826***	0.1819*	0.1909***	0.5523***
	(0.01288)	(0.09)	(0.01422)	(0.07029)
Number of family members	1.809**	3.284**	6.434***	13.64***
	(0.6675)	(1.11)	(0.7368)	(0.8671)
Reading newspaper	-1.647	2.789	3.203*	8.047***
	(1.429)	(2.839)	(1.578)	(2.217)
Reading magazine	2.693	-6.101*	-0.6387	-9.208***
	(1.557)	(2.726)	-1.719	(2.129)
Listening to broadcast	-0.7122	0.829	-0.8816	0.9333
	(0.9861)	(2.081)	(1.089)	(1.625)
Watching TV	1.18	3.281	-0.05804	3.265*
	(0.928)	(1.673)	(1.024)	(1.307)
Sending text	-1.491	-1.751	-5.4**	-3.947
	(1.596)	(2.741)	(1.762)	(2.14)
Relaxing	1.998	-4.38	4.445***	-3.702*
	(1.149)	(2.315)	(1.268)	(1.808)
Studying	-0.2295	1.693	-1.935	-7.673***
	(1.23)	(2.073)	(1.357)	(1.619)
Social with neighbors	-2.388***	-4.109***	-3.579***	-3.133***
	(0.5272)	(0.9787)	(0.582)	(0.7644)

Note: *p<0.05; **p<0.01; ***p<0.001

Data from CGSS

Table 3.2 Results of Baseline Model-Continued

	Dependent variable:			
	inc		cnc	
	(1)	(2)	(3)	(4)
Social with friends	0.09425	-0.4639	1.441	1.063
	(0.6868)	(1.534)	(0.7582)	(1.198)
Watching movie in cinema	3.759*	4.402	10.33***	10.05***
	(1.701)	(2.662)	(1.877)	(2.079)
Shopping	1.72	-2.484	2.357*	-5.869***
	(0.962)	(2.238)	(1.062)	(1.748)
Attending concert	-3.414*	-3.035	-3.89*	1.168
	(1.459)	(2.522)	(1.611)	(1.97)
Party with family members	-0.8152	1.124	-6.944***	-6.28**
	(1.454)	(2.724)	(1.605)	(2.128)
Party with friends	2.793*	2.262	5.407***	5.774**
	(1.244)	(2.264)	(1.374)	(2.104)
Listening to music at home	0.2849	1.004	3.812***	4.674***
	(0.7795)	(1.55)	(0.8604)	(1.211)
Sport exercise	-2.727***	-6.044***	-0.05523	0.5264
	(0.7152)	(1.535)	(0.7895)	(1.199)
Watching sports games on site	3.053	3.456	7.133***	7.472**
	(1.938)	(2.929)	(2.14)	(2.288)
Doing manual work	0.3345	0.1317	0.08662	-3.236
	(1.14)	(2.121)	(1.259)	(1.657)
Surfing Internet	1.508	2.795	0.7764	2.914
	(1.606)	(3.419)	(1.773)	(2.671)
Books read in last 12 months	0.03039	-0.04615	-0.07418	-0.135
	(0.05787)	(0.08279)	(0.06388)	(0.06466)
E-books read in last 12 months	-0.08486	-0.06505	0.006876	0.02552
	(0.06899)	(0.07872)	(0.07616)	(0.06148)
Newspaper and magazine read in last 12 months	-0.003186	0.01121	-0.01082	0.005813
	(0.01136)	(0.02706)	(0.01254)	(0.02113)
Read online information in last 12 months	-0.01448	-0.01601	-0.02712*	-0.01376
	(0.01212)	(0.01444)	(0.01338)	(0.01128)
Having a cell phone	-7.131*	-20.84	-22.79***	30.16*
	(3.439)	(19.69)	(3.796)	(15.38)
Being online in last 6 months	-7.927	7.76	-26.73***	-33.57*
	(4.941)	(17.72)	(5.454)	(13.84)
Other family member being online in last 6 months	1.299	2.169	6.223*	22.09***
	(2.633)	(7.34)	(2.907)	(5.733)
Wechat Pay in last 6 months	-4.614	2.618	-9.686*	-4.933
	(4.077)	(8.758)	(4.5)	(6.84)
Alipay in last 6 months	3.622	0.1477	-2.711	-0.03335
	(3.897)	(6.158)	(4.302)	(4.809)
Constant	190.4***	217***	223.5***	18.47
	(20.46)	(45.3)	(22.59)	(35.38)

Note: *p<0.05; **p<0.01; ***p<0.001

Data from CGSS

Table 3.2-Continued

	Dependent variable:			
	inc		cnc	
	(1)	(2)	(3)	(4)
Observations	11595	2519	11595	2519
R^2	0.0604	0.06594	0.3022	0.4171
Adjusted R^2	0.05698	0.0501	0.2997	0.4072
Residual standard error	104.2 (df=11552)	88.6 (df=2476)	115.1 (df=11552)	69.2 (df=2476)
F-statistic	17.68*** (df=42;11552)	4.162 (df=42;2476)	119.1*** (df=42;11552)	42.18 (df=42;2476)
White statistic	96.7	74.4	149***	340***
Breusch-Pagan statistic	67.008*** (df=42)	40.365 (df=42)	112.34*** (df=42)	217.97*** (df=42)
Note:				*p<0.05; **p<0.01; ***p<0.001
				Data from CGSS

variable is the ideal number of children too, but the dataset is a smaller one that is filtered by all observations with birth dates after 1981, namely the young people. In the third model, the dependent variable is the current number of children, and the dataset includes all observations. In the last model, the dependent variable is the current number of children, and the dataset is the same as the second model.

OLS estimation is applied in the baseline model. The White test and Breusch-Pagan Test are implemented on the baseline model to show whether there is heteroscedasticity in the model. If the results of White test or Breusch-Pagan Test are significant, it indicates that there is significant heteroscedasticity in the baseline model. If there is heteroscedasticity in the model then the assumptions of Gauss-Markov theorem cannot be satisfied, which shows that OLS may not be the optimal unbiased estimation, and we have reason to use other models.

There is a problem with the baseline model, that is, there are 42 independent variables, which may make the baseline model have a high variance. We can improve the performance of the model by reducing the variance of the model. There are two ways to reduce the variance of the model, one is to shrink the coefficients towards zero, the other is to reduce the

number of independent variables. Both methods will increase the bias of the model. Ridge belongs to the first method, and Lasso model belongs to the second method. An issue encountered in research is to select more relevant explanatory variables from a series of variables. Therefore, after estimating the baseline model, shrinkage regression methods are used to modify the model. There are two shrinkage methods: Ridge regression and Lasso regression.

The coefficients given by Ridge model are obtained from the formulation:

$$\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{i=1}^{\mathrm{p}}\beta_j{}^2=\mathrm{RSS}+\lambda\sum_{i=1}^{\mathrm{p}}\beta_j{}^2 \qquad (3.3)$$

If the terms associated with λ are removed from the formulation above, then the formulation will be reduced to RSS that was applied in least squares fitting procedure. The advantage of the Ridge model is that shrinking the coefficients towards 0 can significantly reduce their variance (Hoerl and Kennard, 2000). Ridge regression's advantage over least squares is rooted in the bias variance trade-off. As λ increases, the flexibility of the Ridge regression fit decreases, leading to decreased variance but increased bias. Meanwhile the coefficient decreases and shrinks to 0. Here the λ is chosen by the value that minimizes MSE. Finally, coefficients are biased, but the variance is reduced. Here we have chosen to implement the function over a grid of values ranging from $\lambda = 10^{10}$ to $\lambda = 10^{-2}$, essentially covering the full range of scenarios from the null model containing only the intercept, to the least squares fit. The Ridge model is not a variable selection method, which means all variables in the baseline model will be retained in Ridge, which allows us to compare the results of the two models.

The coefficient given by Lasso model is obtained from the formulation:

$$\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{i=1}^{\mathrm{p}}|\beta_j|=\mathrm{RSS}+\lambda\sum_{i=1}^{\mathrm{p}}|\beta_j| \qquad (3.4)$$

The Lasso model looks very similar to the Ridge model, but the difference is that the term associated with λ is no longer square but absolute value (Tibshirani and Robert, 1996). In the Lasso model, when λ increases the coefficients will be shrinking to 0 as well as in the Ridge model. In this process, the bias increases and the variance decreases. We have chosen to implement the function over a grid of values ranging from λ = 1010 to λ = 10−2 , the same we do in the Ridge model.

The difference between Lasso regression and Ridge regression is that in Ridge regression, as λ increases, all coefficients will shrink to 0, but no coefficient will be exactly equal to 0. In Lasso regression, as λ increases, some coefficients will be exactly equal to 0. Neither Ridge regression nor Lasso will universally dominate the other. In general, we might expect Ridge regression to perform better when the response is a function of many predictors, all with coefficients of roughly equal size. We might expect Lasso to perform better in a setting where a relatively small number of predictors have substantial coefficients, and the remaining predictors have coefficients that are very small or equal to zero. Therefore, Lasso regression is a variable selection method, hence results in Lasso models are easier to interpret. Thus, Lasso gives us an option that compares its result with the baseline model.

4. Empirical results

In the baseline model firstly, we observed that the differences between provinces have a significant impact on the ideal number of children in models 1, 3 and 4. As a control variable, we do not know more details of the impact of provincial differences. If we want to understand the impact differences between specific provinces, a possible improvement method is to set different provinces as dummy variables.

In model 4, we observed statistically significant gender differences with large negative effects, which shows that women born after 1981 in China have more children. Ethnicity difference, which is also a control variable, is statistically insignificant in model 1, 2 and 3. Therefore, we cannot observe the differences between specific ethnic groups. If we want to understand the impact differences between specific ethnicity groups, a better improvement is to set different ethnicity as dummy variables.

In models 1, 2 and 3, the frequency of religious activities is statistically significant, and the coefficients are positive, which means that the frequency of religious activities has a positive impact on the number of ideal children. This result is consistent with the classic conclusion of the past studies, that is, those with religious conservative attitudes tend to have more children. Religious activities are not significant in model 4, indicating that religion may have less impact on Chinese young people.

In model 1, 3 and 4, education is statistically significant and has a negative impact. This is consistent with the classical conclusion that the improvement of education level makes people have fewer children. In model 1 and 3, income is not statistically significant, but in model 2 and 4, income is significant. One explanation for this is that young Chinese are more sensitive to income factors when considering having children than all Chinese. This result explains the difference between Malthusian framework and non-Malthusian framework to some extent. In the Malthusian framework, when income increases, people always tend to have more children, but it does not occur in non-Malthusian framework. It is understandable that the two theories coexist for a country in the transition period such as China and showing that income is not statistically significant to the ideal number of children. The insignificance of income to the ideal number of children also gives us another hint, that is, simple financial subsidies are not likely to be effective to encourage fertility.

The house size is positively significant in model 1 and 3 and the coefficient is positive,

while it is not statistically significant in model 2 and 4, which means that people tend to have more children when they have larger houses, or when housing conditions improve, people tend to have more children. However, Chinese young people are less concerned about housing conditions.

In model 3, height is statistically significant, but surprisingly, it has a negative effect on both the ideal number of children and the current number of children, which is different from common sense. In contrast, weight only showed statistical significance in model 4, and only had a small positive effect on the current number of children. It is also surprising that health, that is, the respondents' evaluation of their own health, is significant in model 3, but has a negative effect on the current number of children, which is inconsistent with common sense.

As another control variable, the difference between urban and rural areas shows statistical significance in models 1, 2 and 3. We can speculate that rural areas have a greater positive effect on the number of children than cities from past studies, but we cannot observe it here. Working hours[7] are only statistically significant for the ideal number of children in model 1 and insignificant in rest models.

The maximum unemployment months are statistically significant in all the four models, and surprisingly the coefficients are positive rather than negative. Literally, the longer the time of unemployment, the more children people tend to have, which seems puzzling. Conversely, can we say that long-term employment has a negative effect on the number of children people want to have? We hope to further explain this result. The last variable of basic information, the number of family members, is statistically significant in all four models. The independent variable of the number of family members has an obvious collinear relationship with the current number of children, so it is not surprising that this variable

[7] Working hours may be an endogenous variable, but here we treat it as an exogenous variable.

is statistically significant.

In the section of individual activity, firstly, “the frequency of reading newspapers in the last year” is positively significant in model 3 and 4, which shows that those who read more newspapers tend to have more children. “The frequency of reading magazines in the last year” is negatively significant in model 2 and 4, which shows that among Chinese young people who read magazines more frequently tend to have fewer children. “The frequency of listening to the radio last year” is not statistically significant in the four baseline models. “Frequency of watching TV in the last year” is positively significant in model 4, which shows that young people in China who watch more TV tend to have more children. “Frequency of sending text in last year” is negatively significant in model 3, which shows that people who send more text messages tend to have fewer children.

In model 3 and 4, “frequency of relaxation in last year” is statistically significant, but the sign of the two coefficients from the two models are opposite, which indicates that the specific impact of relaxing on the number of children needs to be further studied. “Frequency of study in the last year” is statistically significant in model 4 and has a relatively large negative effect, which shows that young people in China who spend more time on studying tend to have fewer children. Considering the possible collinearity between learning and education level, this result is understandable.

In all four models, “the frequency of socializing with neighbors in the last year” is statistically significant. The negative coefficient here shows that people who spend more time on socializing with neighbors tend to have more children, which is consistent with the hypothesis of this paper.

It is also a direct social activity, but the “frequency of social with friends in the last year” does not show statistical significance, which is inconsistent with the hypothesis of this paper.

“The frequency of watching movies in cinema in the last year” has a significant positive effect on the dependent variable in model 1, 2 and 4, which means that people who engage in more of this activity tend to have fewer children. Considering that the activity is considered as a non-social activity, this result supports the hypothesis of this paper.

“Frequency of shopping in last year” has a significant impact on the dependent variable in model 3 and model 4, but the sign of the coefficient is opposite, which shows that the impact of shopping activity on the number of children is difficult to determine.

“Frequency of attending concerts in last year” has a significant negative effect on the dependent variable in model 1 and model 3, which shows that in the sample of all Chinese people, participating in more concerts tends to have fewer children.

“Frequency of party with family members in the last year” has a significant negative effect on dependent variables in model 3 and 4, which shows that people who spend more time on family parties tend to have fewer children.

“Frequency of party with friends last year” has a significant positive effect on dependent variables in model 3 and 4, which shows that people who spend more time on party with friends tend to have more children. Comparing the two we can infer that this is a further confirmation of the hypothesis of this paper.

“Frequency of listening to music at home in last year” has a significant positive effect on dependent variables in model 3 and 4, which shows that those who spend more time on listening to music at home tend to have less children, which is another confirmation of the hypothesis of this paper.

“The frequency of sports exercise in the last year” has a significant negative effect on the ideal number of children in models 1 and 2, which shows that more sport activities tend to have more children.

“Watching sports games on site last year” has a significant positive effect in model 3

and 4, which shows that more people who watch sports games on site have fewer children.

“Reading online information in the last 12 months” has a statistically significant but small negative effect in model 3.

“Having a mobile phone” is statistically significant in model 1, 2 and 4, but the coefficient is positive in model 1 and 3 and negative in model 4, so we cannot judge the effect of this variable here.

“Surfing Internet in last 6 months” is statistically significant in model 3 and 4, and both have large negative coefficients, indicating that surfing the Internet has a large negative effect on the number of children.

However, “other family members surfed the Internet in the last 6 months” is positively significant in model 3 and 4. This result is contradictory to the result of the previous variable, “surfing Internet in last 6 months”.

Finally, “wechat pay in the last 6 months” has a significant negative effect in model 3, indicating that using more mobile phone payment may have a negative impact on the number of children. This result is consistent with the result of the variable “having a mobile phone”. “Wechatpay in the last 6 months” shows significance, while "Alipay in the past 6 months” does not show significance. The two mobile apps with similar utility and mutual substitutes show different results, which need to be explored further.

Figure 3.1 shows how the value of λ is selected in each of the Ridge models. Among the four Ridge models, the coefficients in model 2 are very small and close to 0, because a relatively large coefficient λ is selected in the model. When we ignore model 2 and first interpret the results of models 1, 3 and 4, we find that the coefficients of those variables that show statistical significance in the baseline model still retain relatively large coefficients in the Ridge results.

Table 3.3 shows the results of Ridge models. In model 1, if we rank all the absolute

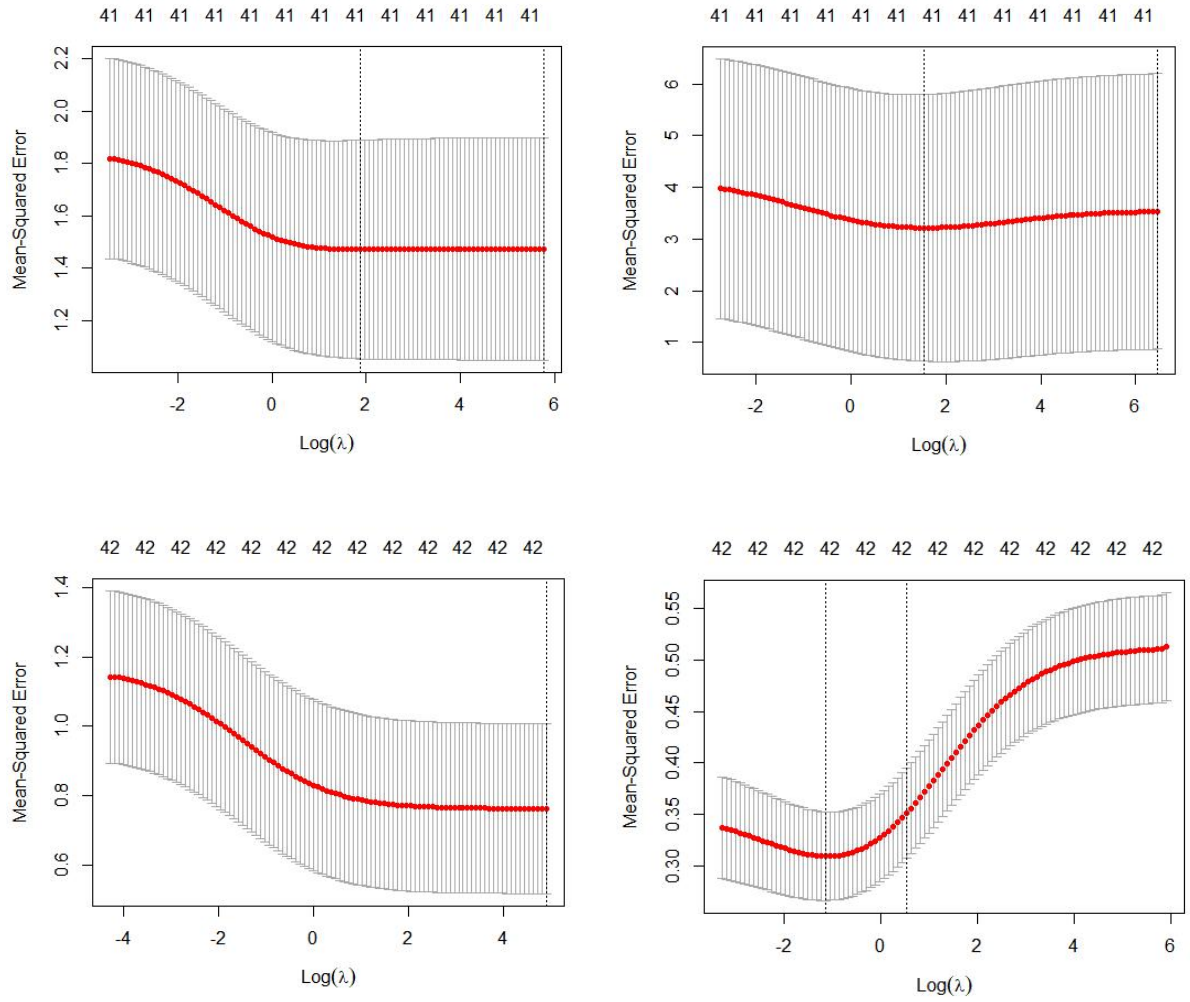

Figure 3.1 MSE-Log(λ) in Ridge Models

values of the coefficients then among the top 10 variables there is only one variable having a great positive impact on the ideal number of children, that is “frequency of religious activities”. These variables have a great negative impact on ideal number of children: “watching sports games on site”, “watching movie in cinema”, “being online in last 6 months”, “wechat pay in last 6 months”, “have a cell phone”, “alipay in last 6 months”, “other family member being online in last 6 months” and “sending text”. All the results are consistent with the hypothesis of social activity except the activity “watching sports games on site”.

In model 3, we rank all the absolute values of the coefficients to obtain among the top 10 variables. All the top 10 variables have a negative impact on the current number of

Table 3.3 Results of Ridge Model

	Dependent variable:			
	inc		cnc	
	(1)	(2)	(3)	(4)
Province	0.001	6.23E-39	0.002	2.137E-04
Gender	-5.040E-4	4.54E-38	-0.014	-0.126
Ethnicity	0.003	2.73E-38	0.003	-0.014
Religion	0.008	7.85E-38	0.016	0.015
Education	-0.004	-2.32E-38	-0.019	-0.028
Income	-6.497E-09	2.31E-43	-1.407E-08	8.823E-07
House size	9.449E-05	7.51E-40	2.392E-04	2.668E-04
Height	-3.127E-04	-1.85E-39	-0.002	-5.752E-04
Weight	-6.752E-05	9.86E-40	-4.050E-04	0.001
Health	-0.004	-4.01E-38	-0.034	-0.017
Urbanization	-0.012	-7.18E-38	-0.028	-0.014
Working hours per week	1.883E-07	1.47E-39	-5.19E-04	9.418E-04
Unemployment months	6.219E-05	2.61E-39	4.829E-04	0.004
Number of family members	0.003	5.73E-38	0.014	0.102
Reading newspaper	-0.004	2.91E-39	-0.004	0.037
Reading magazine	-0.004	-4.71E-38	-0.020	-0.056
Listening to broadcast	-0.003	6.01E-39	-0.006	0.006
Watching TV	0.004	5.34E-38	0.010	0.035
Sending text	-0.008	-7.61E-38	-0.032	-0.042
Relaxing	0.003	-4.09E-38	0.012	-0.023
Studying	-0.006	-3.96E-38	-0.029	-0.072
Social with neighbors	-0.005	-5.79E-38	-0.015	-0.029
Social with friends	6.040E-04	4.68E-40	0.008	0.014
Watching movie in cinema	0.013	9.78E-38	0.057	0.084
Shopping	0.004	-5.76E-39	0.013	-0.045
Attending concert	0.001	5.73E-39	0.014	0.023
Party with family members	0.002	-9.08E-39	-5.091E-04	-0.051
Party with friends	0.002	4.70E-38	0.026	0.049
Listening to music at home	0.004	3.50E-38	0.022	0.043
Sport exercise	-3.852E-04	-2.13E-38	0.009	0.015
Watching sports games on site	0.007	1.66E-38	0.035	0.063
Doing manual work	5.740E-04	-2.45E-38	0.002	-0.031
Surfing Internet	0.007	7.36E-38	0.028	0.031
Books read in last 12 months	-1.064E-04	-1.57E-39	-5.439E-04	-0.001
E-books read in last 12 months	-1.924E-04	-1.60E-39	-4.533E-04	-7.511E-05
Newspaper and magazine read in last 12 months	-2.700E-05	-6.65E-41	-2.915E-05	9.349E-05
Read online information in last 12 months	-9.736E-05	-4.73E-40	-3.497E-04	-1.503E-04
Having a cell phone	-0.022	-3.66E-37	-0.107	0.195
Being online in last 6 months	-0.028	-2.66E-37	-0.109	-0.211
Other family member being online in last 6 months	-0.010	-2.49E-38	-0.034	0.175
Wechat Pay in last 6 months	-0.023	-1.66E-37	-0.098	-0.055
Alipay in last 6 months	-0.021	-1.47E-37	-0.096	-0.038
Constant	1.961	1.724	1.740	0.433
Observations	11595	2519	11595	2519

Data from CGSS

children. These variables are: “being online in last 6 months”, “have a cell phone”, “wechat pay in last 6 months”, “watching movie in cinema”, “watching sports games on site”, “other family member being online in last 6 months”, “health”, “sending text”, “studying”. All the results are consistent with the hypothesis of social activity except the activity “watching sports games on site”, same as the result in model 1.

In model 4, if we rank all the absolute values of the coefficients then among the top 10 variables, the following variables have a great positive impact on the current number of children: “having a cell phone”, “other family members being online in the last 6 months” and “number of family members”. The first two results are against the results in model 1 and 3, so as it is against the hypothesis in the paper. The top 10 variables that have a great negative impact on the current number of children are “being online in the last 6 months”, “gender”, “watching movies in cinema”, “studying”, “watching sports games on site”, “reading magazines”, “wechat pay in the last 6 months”. All the results are consistent with the hypothesis of social activity except the activity “watching sports games on site”, same as the result in model 1 and 3. Finally, the difference between urbanization has a large coefficient in the results of model 1, 3 and 4, and we do not need to explain too much because this is just a control variable.

Figure 3.2 shows that in each Lasso model how the value of λ is selected. In model 1, 3 and 4, we can clearly see the existence of the minimum MSE. In model 2, the minimum MSE is on the very right side of the horizontal axis, so the value of λ is so large that all the coefficients of model 2 are equal to 0 resulting in the model becoming a null model. As a result, Lasso model 1 retains 4 variables, model 2 does not retain any variables, model 3 retains 6 variables, and model 4 retains 26 variables at most.

Figure 3.3 shows how these variables are added to the null model in turn. In other words, Lasso gives us 3 shrinkage models with higher bias and lower variance compared to

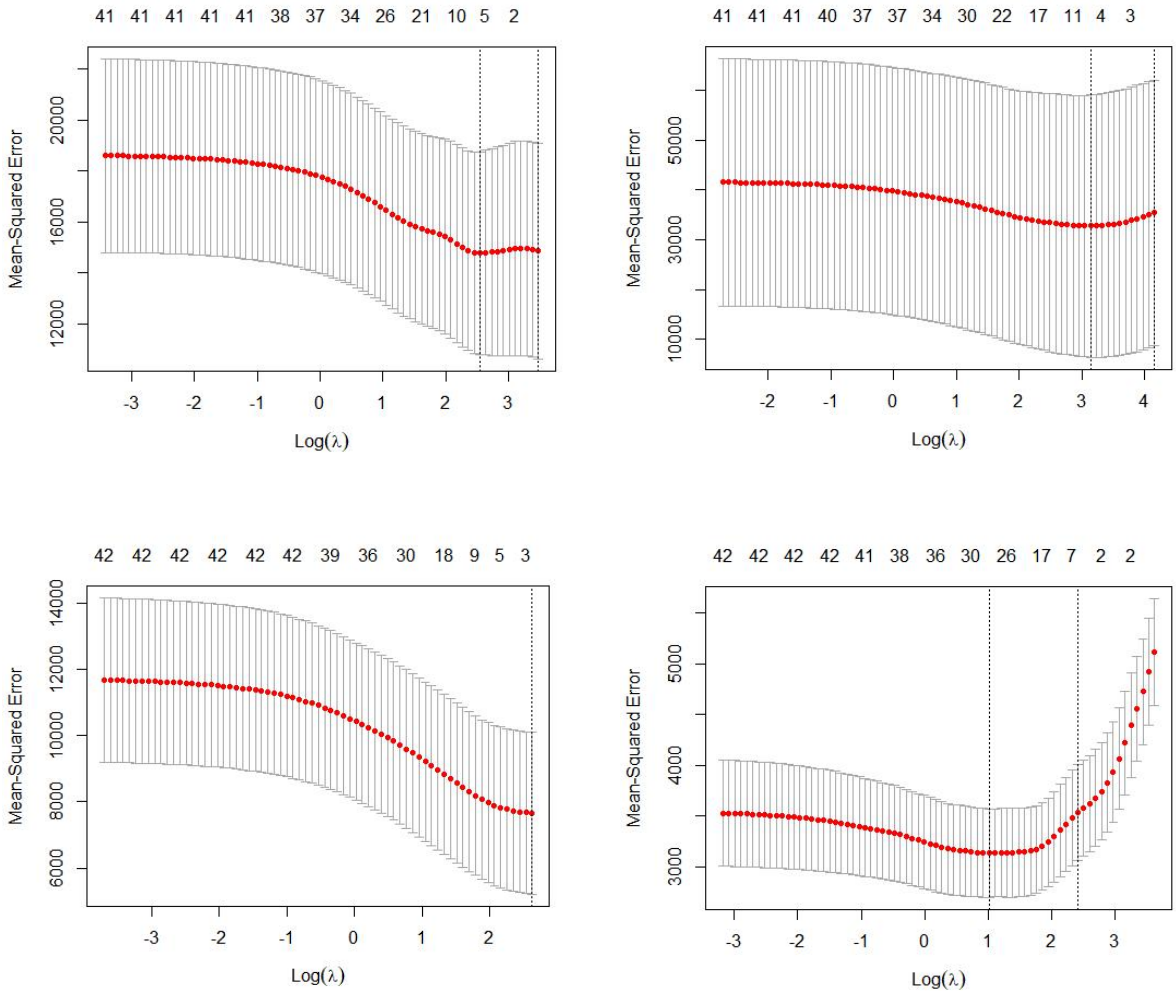

Figure 3.2 MSE-Log(λ) in Lasso Models

the baseline model.

Table 3.4 shows the result of Lasso model. In the Lasso model, education is retained in all three models, and the coefficients are negative, which shows that the improvement of education has a negative effect on the number of children. Similarly, the "frequency of sending text in last year" is retained in all three models, and the coefficient is negative in all three models, indicating that those who send more text tend to have fewer children. The "frequency of watching movies in cinema in the last year", which is significant in the baseline model, is retained in Lasso models 3 and 4, and still has positive coefficient, which is consistent with the results obtained in the baseline model. "Surfing Internet in last year" is

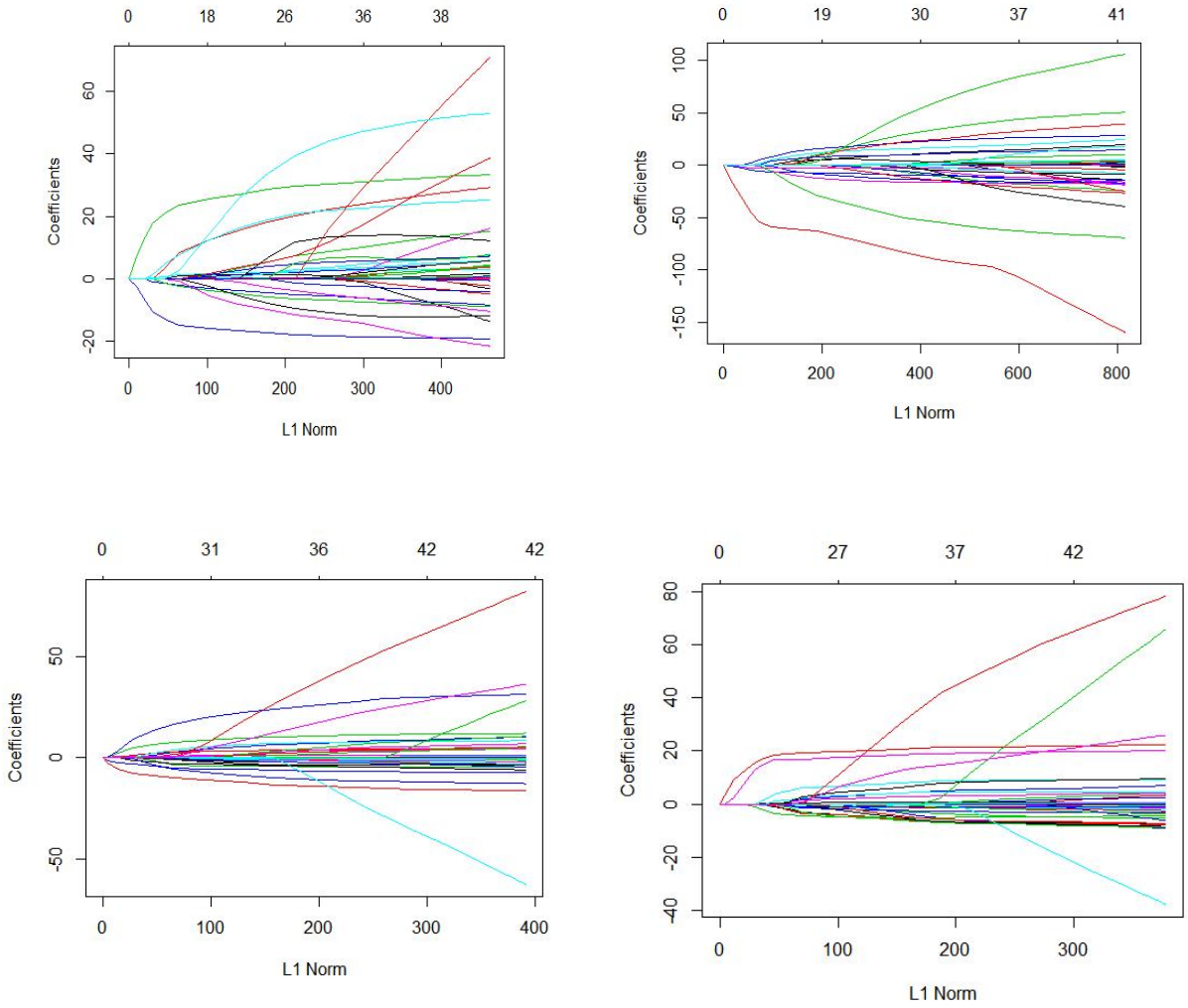

Figure 3.3 Coefficients in Lasso Models

retained in models 1, 3 and 4, which is consistent with the results obtained in the baseline model, and the coefficient of this variable is negative, indicating that surfing Internet has a negative impact on the number of children. Consistent with this result, the variable “being online in the last 6 months” is retained in model 1 and 3, and the coefficient is negative. “Wechat Pay in the last 6 months” is retained in model 3 and has a large negative effect, which is consistent with the results in the baseline model.

5. Conclusion

The empirical results are essentially consistent with the hypothesis. The empirical

Table 3.4 Results of Lasso Model

	Dependent variable:			
	inc		cnc	
	(1)	(2)	(3)	(4)
Province				
Gender				-12.624
Ethnicity				
Religion				
Education	-0.493		-5.553	-3.715
Income				9.152E-05
House size				9.500E-05
Height				
Weight				0.089
Health				
Urbanization				
Working hours per week				
Unemployment months				0.465
Number of family members				13.268
Reading newspaper				0.403
Reading magazine				-2.006
Listening to broadcast				
Watching TV				2.221
Sending text	-2.382		-9.765	-4.112
Relaxing				-0.036
Studying				-9.221
Social with neighbors				-2.345
Social with friends				
Watching movie in cinema			1.689	9.156
Shopping				-2.636
Attending concert				
Party with family members				-2.838
Party with friends				4.366
Listening to music at home				4.319
Sport exercise				
Watching sports games on site				4.103
Doing manual work				-1.395
Surfing Internet	0.608		1.384	1.894
Books read in last 12 months				-0.065
E-books read in last 12 months				
Newspaper and magazine read in last 12 months				
Read online information in last 12 months				-0.003
Having a cell phone				-11.198
Being online in last 6 months	-0.008		-13.845	
Other family member being online in last 6 months				9.409
Wechat Pay in last 6 months			-5.777	
Alipay in last 6 months				
Constant	204.330	172.449	225.725	46.422
Observations	11595	2519	11595	2519

Data from CGSS

results show that some “traditional” factors that have been emphasized to have a significant impact on the number of children in the past, such as education, religion, provincial differences, urban-rural differences, are still statistically significant in this study.

More importantly, we find that some individual activities do have an impact on the number of children that cannot be ignored, and the impact from those variables are even no less than that of some traditional factors. In the variables of social activities that we defined, except “watching sports games on site”, most variables have a significant positive impact on the number of children. In the variables of non-social activities, many variables have a significant negative impact on the number of children. Especially in terms of Internet activities, the results are that these activities, such as having a cell phone or surfing the Internet, have a great negative impact on the number of children.

The models used in this paper are linear models, so we cannot correctly know whether the independent variables have high-dimensional impact on the dependent variable, thus we must introduce high-dimensional models into future research. Also, the models do not consider the problem of endogeneity.

As we have observed from the baseline model, the impact of income on both the ideal number of children and the current number of children is not as significant as other factors. Therefore, policymakers may not expect to simply adopt the method of financial subsidies or incentives to encourage fertility, although this method may be simple and easy to implement. However, this does not mean that policymakers can make policies that simply influence other variables to promote fertility rate, such as gender differences, religious differences and educational inequality.

We must point out that the activity patterns that affect the number of children is part of human modern life, and this new lifestyle is the result of economic development. Even if we are faced with the risk of low fertility trap or population shrinkage in the future, it is

impractical for us to go back to the pre-modern society, such as the introduction of policies to reduce the level of education or reduce the number of mobile phones to increase the fertility rate. No country has completely jumped out of the low fertility trap today. We cannot say that the current policies to encourage fertility implemented by some countries are ineffective, but we may try more innovative methods, such as increasing diversified public services, creating new social space, and most importantly, to find if there is a new way which is suitable for this era that is full of entertainment, Internet and low social networking.

References

Bauer, Gerrit, and Thorsten Kneip. 2013. "Fertility from a Couple Perspective: A Test of Competing Decision Rules on Proceptive Behaviour." *European Sociological Review*, Volume 29, Issue 3, June 2013, Pages 535–548.

Bernardi, Laura, and Andreas Klärner. 2014. "Social Networks and Fertility." *Demographic Research*, Volume 30, Article 22.

Bongaarts, John. 2003. "Completing the Fertility Transition in the Developing World: The Role of Educational Differences and Fertility Preferences." *Population Studies*, Vol. 57, No. 3, 2003, pp. 321-336.

Derose, Laurie F. and Alex C. Ezeh. 2007. "Men's influence on the onset and progress of fertility decline in Ghana, 1988–98." *A Journal of Demography*, Volume 59, 2005 - Issue 2.

Dong, Yige. 2021. "The crisis of social reproduction and 'Made-in-China' feminism." *Soundings*, Number 79, Winter 2021, pp. 10-23(14).

Galor, Oded, and David N. Weil. 1993. "The Gender Gap, Fertility and Growth." *NBER Working Paper Series* #4550.

Galor, Oded, and David N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *The American Economic Review*, Vol. 90, No. 4 (Sep. 2000), pp. 806-828.

Hoerl, Arthur E., and Robert W. Kennard. 2000. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics*, Feb. 2000, Vol. 42, No. 1, Special 40th Anniversary Issue (Feb. 2000), pp. 80-86

Kaufman, Gayle, and Eva Bernhardt. 2012. "His and Her Job: What Matters Most for Fertility Plans and Actual Childbearing?" *Family Relations*, Volume 61, Issue 4.

Kinsella, Sharon. 1998. “Japanese Subculture in the 1990s: Otaku and the Amateur Manga Movement.” *The Journal of Japanese Studies*, Summer, 1998, Vol. 24, No. 2 (Summer, 1998), pp. 289-316

Kunz, Phillip R. 1965. “The Relation of Income and Fertility.” *Journal of Marriage and Family*, Nov. 1965, Vol. 27, No. 4 (Nov. 1965), pp. 509-513.

Martin, Teresa Castro. 1995. “Women’s Education and Fertility: Results from 26 Demographic and Health Surveys.” *Studies in Family Planning*, Jul. - Aug. 1995, Vol. 26, No. 4 (Jul. - Aug. 1995), pp. 187-202.

Koula Mellos. 1988. “Neo-Malthusian Theory.” *Perspectives on Ecology*, pp 15–42.

Muradyan, Olena, and Daria Yashkina. 2020. “A Study of Solo-Livers in Eastern Asia: An Eye on Japan, China and South Korea” *Sociological Studios*, 2(17), 76–84.DOI: 10.29038/2306-3971-2020-02-76-84

Tibshirani, Robert, 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society*. Series B (Methodological), 1996, Vol. 58, No.1 (1996), pp. 267-288

Wang, Feng, Baochang Gu and Yong Cai. 2016. “The End of China's One-Child Policy.” *Studies in Family Planning*, Volume47, Issue1, Pages 83-86.

Yu, Binbin. 2021. “Ecological effects of new-type urbanization in China.” *Renewable and Sustainable Energy Reviews*,Volume 135, January 2021, 110239.

CONCLUSION

The dissertation provides a nuanced analysis of the factors contributing to China's declining fertility rates, utilizing a blend of econometric methods and machine learning tools. It approaches the issue from three distinct microeconomic perspectives, offering a multi-faceted understanding of the dynamics at play.

One of the key insights from the research is the complex interplay between education levels and fertility rates. It reveals that the education level of family members, particularly the respondent and their spouse, is inversely related to the number of children, with a higher education level leading to fewer children. However, this effect is less pronounced when considering the education level of the parents.

The research goes beyond the traditional view, suggesting that not just the absolute education level, but the difference in education levels between family members, significantly impacts fertility rates. This difference, indicative of a sort of "comparative advantage" within the family, leads to a division of labor that could enhance family income and, subsequently, fertility rates. Interestingly, the study also uncovers that the impact of educational differences varies by gender. In households where the respondent is female, the disparity in education levels between her and her spouse has a smaller effect on the number of children.

Additionally, the dissertation explores the relationship between income and fertility rates. It challenges the conventional wisdom of a simple negative correlation, suggesting a more nuanced relationship. The positive correlation holds true primarily for those people who are engaged in agricultural jobs, whereas no significant link is found for other professions. This finding underscores the role of economic factors beyond mere income, like house size, in influencing fertility decisions.

The research also delves into how modern lifestyles, shaped by the proliferation of non-social activities like internet use, have led to a decline in fertility rates. These activities,

by reducing time for social interaction, indirectly affect decisions regarding marriage and childbirth.

Based on these findings, it becomes apparent that addressing China's low fertility rates requires a multifaceted approach. Policies need to be cognizant of the diverse factors influencing fertility decisions, from education and income to lifestyle choices. For instance, educational policies could be tailored to acknowledge the impact of educational disparities within families on fertility rates. Support systems for families, especially those with lower overall education levels, could help balance educational, professional, and familial responsibilities. Additionally, encouraging a balance between social and non-social activities could mitigate the negative impact of modern lifestyles on fertility rates.

Future research could further elucidate these relationships. Exploring how family divisions of labor influence household welfare and fertility, examining the role of digital technologies in lifestyle choices, and investigating the varying impacts of different income types on fertility rates could provide deeper insights. Such research would not only enhance our understanding of the factors affecting fertility rates but also guide effective policymaking to address demographic challenges.