The impact of educational level on the sex ratio at birth in China

Qifen Yuan

Master Thesis for the Master of Philosophy in Economics

UNIVERSITETET I OSLO

May 2014

Preface

Finally I finished this thesis and it is time to say goodbye to the wonderful time I have spent in University of Oslo. It is also the time to say thanks to people who are special for me.

My deepest gratitude goes first and foremost to my supervisor Nico Keilman. He led me to this interesting topic. He always gave me exciting inspirations and provided insightful suggestions in the empirical and theoretical parts. I have learnt so much from him. I am very grateful for his patience and kindness to me. He is the person I know with the greatest mind and the warmest heart. He gave me great courage to move on with the thesis. Without him, I would have never finished the thesis on time.

I would also like to thank Erik Biørn, Harald Goldstein and Ragnar Nymoen for their wonderful lectures on Econometrics and Statistics. And also thanks to Nils Christian Framstad and Bjorn Dapi for making Mathematics so interesting. Thanks to Aanund Hylland and Geir B. Asheim for the inspiring lectures on Microeconomics and Game theories. Also thanks to Asbjørn Rødseth for his challenging but interesting lectures on Macroeconomic theories and International Macroeconomics.

I am also very grateful to professor Yuan Cheng and professor Li Zhang from Fudan University, for their supports to me.

I am also thankful for all the friends who shared knowledge and experience during my study, who made the days full of memories.

Finally, I dedicate this thesis to my parents, who gave me so much love and encouragement.

May 2014

Qifen Yuan

Summary

Birth masculinization in China since 1980s has drawn much attentions of scholars as well as politicians. As in many other Asian countries, there have been consistently more boys born than girls in China. Given the huge population in China, an unusually high Sex Ratio at Birth (SRB) defined as the number of baby boys per 100 baby girls born during a certain period, implies a great change in the future sex structure, and thus has long-term effects on almost every aspect of the society.

The earlier studies found that education has a positive effect on SRB because people with higher level of education have better access to sex-selective technology and thus tend to have higher level of SRB than the less educated. However, now the sex-selective technology is widely available and not costly to get. So whether or not to perform sex selection on the child is based on how much boys are preferred over girls. In general, people with higher education will have open mind and not be restrained by the traditional preference for son. Therefore, education should have more discernable effects on decreasing SRB.

In this thesis, we study the determination of the birth masculinity with special focus on the effects of education. First, provincial and county level from the 2010 census data in China are used to analyze the determinants for the great regional differentials of SRB. We perform correlation and regression analyses. Multiple regression models are established to explain the determinants of the overall SRB as well as the SRB for the first two orders. To indicate the education level in the region, we calculate the average numbers of years of schooling of the population in the region, for the total population and for females and males. We also calculate the relative education level of female and male. Then we include the overall education and the relative sex-specific education levels into our regression models, together with other explanatory variables suggested by the earlier findings.

In view of the problem that we were using the aggregate data to make inferences about relationships at the individual level, the individual level data from China General Socio Survey in 2008 is collected. With these data, we analyze how sex composition of the children is determined

by parents' and households' characteristics. We generate a variable called "degree of son preference" base on the sex composition of the children in the family, assigning higher value to families with more sons than daughters. Since the dependent variable "degree of son preference" is a categorical variable, we established the General Ordered Response models and perform the Generalized Ordered Logit Estimations. The education variables we use here are the wife's/mother's and the husbands'/father's education, indicated as categorical variables. They enter the General Ordered Response models together with other explanatory variables.

From the multiple regressions, we find that education is an important determinant for SRB differentials across regions. We find that provinces with generally higher level of education tend to have lower SRB for the second births. We find female education has strong negative effect on SRB. Both provincial data and county-level data suggest that regions where women are less deprived in education are associated with lower SRB. It works especially significant on decrease the excess male births among the second children. So improvement in female education seems able to reduce SRBs substantially. On the other hand, the male advantages in education tend to increase SRBs, which is suggested by both levels of data.

Analyses with the individual level data confirm some of our findings from the aggregate data analyses. We find higher level of wife's education is associated with lower level of son preference, manifested by a lower number of boys among the children in the family. And the husband's education level is likely to reinforce son preference, at least in terms of a higher ratio of sons among the children.

Based on these findings we conclude that in current China, education has started to play a role of reducing people's preference for sons. The improvement in people's education level, especially the female education level, will accelerate the pace for the sex ratio at birth to return to the normal range.

Most part of the data collection and organization is performed using Microsoft Excel 2010. And the correlation and regression analyses are performed using STATA 12.0.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation

Birth masculinization in China since 1980s has drawn much attentions of scholars as well as politicians. As in many other Asian countries, there have been consistently more boys born than girls in China. Given the huge population in China, an unusually high Sex Ratio at Birth (SRB) defined as the number of baby boys per 100 baby girls born during a certain period, implies a great change in the future sex structure, and thus has long-term effects on almost every aspect of the society.

The earlier studies found that education has a positive effect on SRB because people with higher level of education have better access to sex-selective technology and thus tend to have higher level of SRB than the less educated. Other studies point out higher education is associated with lower SRB because education changes people's traditional preference for son. Thus education has a negative effect on SRB. In this thesis, we want to find out how education affects SRB, figuring out the sign and magnitude.

So in this thesis, we main want to answer two questions:

How can education explain the regional disparities of SRB in current China?

To find out the answer, we collect aggregate data on two levels: provincial and county level from the 2010 census data in China(NSB, 2010). There are 31 observations in the provincial data and 1069 observations in the county level data. These two data contain the aggregate variables averaged across the population in that region. We will first examine if there are truly great disparities in SRB across regions. Then we will establish multiple regression models to study the determination of the birth masculinity with special focus on the effects of education. Multiple regression models will be established to explain the determinants of the overall SRB as well as the SRB for the first two orders if they are truly account for much variance in the overall SRB. To indicate the education level in the region, we will calculate the average numbers of years of schooling of the population in the region, for the total population and for females and males. We also calculate the relative education level of female and male. Then we will include the overall education and the relative sex-specific education levels into our regression models, together with other explanatory variables suggested by the earlier findings.

How sex composition of children is determined at individual or household level?

In view of the problem that we were using the aggregate data to make inferences about relationships at the individual level, we want to supplement our aggregate analysis with individual level analysis. Source of the individual data is from China General Socio Survey in 2008. The survey was carried out by Department of Sociology, Renmin University of China & Social Science Division, Hong Kong Science and Technology University, and directed by Dr. Li Lulu & Dr. Bian Yanjie (Yanjie & Lulu, 2008). There are 6000 observations including respondents from both urban areas and rural areas. The data has rich information on individual and household's education, socioeconomic and demographic characteristics. We will generate a variable called "degree of son preference" base on the sex composition of the children in the family, assigning higher value to families with more sons than daughters. Since our dependent variable "degree of son preference" is a categorical variable, we will establish the General Ordered Response models and perform the Generalized Ordered Logit Estimations. The education variables we use here are the wife's/mother's and the husbands'/father's education, indicated as categorical variables. They enter the General Ordered Response models together with other explanatory variables.

Analyzing both aggregate and individual data, we will be able to find the relation between SRB and education without running into biased conclusions that could be resulted in by using just one of them. In this sense, our analyses should be consistent and reliable.

1.2 Literature Review

Birth masculinity is a common social and demographical phenomenon in many Asian countries, such as China, India and South Korea. A great number of studies have been devoted to analyze such large excess of male births, trying to find appropriate explanations for it. The common concept we often use is the sex ratio at birth (SRB). It is computed as the number of male births for 100 female births during a certain time frame (usually a year) within a certain geographical or administrative region. A range from 102 to 107 boys born to 100 girls is considered as normal. While over 110 boys to 100 girls is considered as “abnormal”. When the level reaches 115 or higher, it is usually considered as “worrisome”, especially if it is observed for several successive years. There are a large amount of researches on the determination of the inordinate SRBs in different countries. As summarized by Guilmoto (2009), today’s birth masculinization relies on the following three factors:

(1) The entrenched son preference (“readiness”);

(2) Access to modern sex selection technology (“ability”)

(3) The pressure caused by fertility decline (“squeeze”)

Under these circumstances, parents are willing to, able to and have no alternative but to resort to active sex selection of their offspring, leading to an increase in SRB. This is a very concise and powerful framework, under which various factors affect the level of SRB via one or several of the above channels. Earlier findings show that the highest sex ratios are observed in countries with a combination of son preference, easy access to sex selective technology, and low fertility rate (Banister, 2004; Zhu, Lu, & Hesketh, 2009).

Son preference

In the international scope, gender preferences are diversified across different societies. Son preference is not found everywhere and not dominant in many societies (Fuse, 2010). While high levels of SRB are often observed in countries with traditional preference for sons, suggesting a possible causality between attitudinal son preference and the manifested behavior of son preference.

Son preference has its root in Confucian gender norm: the patrilineal kinship. It was a dominant kinship in East Asia during preindustrial periods. It dictated women/wife's absolute obedience to the men/husband in the family. Under such circumstances, men were highly valued than women and women were extremely marginalized. Women's imperative duty was to bear sons for husband's family, ensuring the continuity of the family line. On the other hand, it also defines the duty of sons in the family, such as they must be filial to parents and support their parents when they grow older.

Some study assert that it is the strong preference for son that drives couples to perform sex-selective abortion(Hesketh, 2011). Chuang and Das Gupta(2007) also implies that the rise in SRB in South Korea can be largely attributed to people's traditional preference for son; and the decline since the mid-1990s in SRB is associated with the reductions in son preference in Korean society. Almost all studies on SRB consider son preference as one of the most important reasons for the abnormally high SRB in East Asia. One often discussed question is how important is the role of son preference plays in determining inordinate SRBs, relative to the roles of other factors. Someone think the effects of son preference on high SRB are more prominent than economic factors (Tan, 2008). Another often discussed question is the relation between son preference and modernization and industrialization. In China and India, SRBs seem to increase alongside with the industrialization and modernization for several decades. Some studies suggest that modernization reduces fertility and makes the latent son preference manifested in behavior (Filmer, Friedman, & Schady, 2009). While others point out modernization has positive effects on reducing son preference and thus can reduce SRBs (Chung & Das Gupta, 2007).

<u>Sex-selective abortion</u>

Many studies recognize the role of sex selective abortion plays on occurrence of inordinate SRBs. Before access to sex-selective technology, a usual way to reduce number of girls in the family is by infanticide or neglect of health care for girls, i.e. by great female-biased infant and child mortality. Since legislation against such behavior gets more severe nowadays, female infanticide is very rare. Parents who want to manipulate the sex of child would usually resort to ultrasonographic sex determination and then perform abortion if it diagnoses a female fetus. Therefore, it is an important reason for high SRBs.

Some studies point out that sex selective abortion accounts for almost all the excess male births in China (Zhu et al., 2009). They prove that excess births of boys are not because of under-registration of female births, but due to wide use of sex selective abortion. Attan é (2009) also implies that sex-selective abortion leads directly to the increase in SRB.

China's family planning policy

Another relevant factor is China's family policy. The main mechanism is the policy lower the fertility, people with son preference are "pushed" to use gender-selection technology.

In fact, there are great variations in China's fertility policy across provinces. In urban areas or among people with a nonagricultural household registration status, the majority is under the one-child policy, covering 35.4% of China's population. The "1.5 child policy" is referred to the rule that second child is allowed after certain years of spacing if the first child is a girl. This policy is implemented in the rural areas of 19 provinces, covering about 53.6% of China's population. Only 6 provinces allow two children for rural couples unconditional on the sex of the first child, and this category covers 9.7% of total population. Besides, people of ethnic groups are usually allowed to have two or more children. Those who are only children themselves are also allowed to have two children. (Gu, Wang, Guo, & Zhang, 2007)

China's family planning policy has profoundly changed people's fertility behavior. It not only reduces the total fertility rate to a fairy low level, but also has induced high SRB levels in many provinces. Ebenstein (2010) asserts that there is a causally link between the "missing girls" and the enforcement of the one-child policy. Some study points out that the "1.5-child" policy accounts most for the inordinate SRB, as the highest sex ratio were observed in those provinces (Zhu et al., 2009). Some study even quantifies the contribution of the one-child to the high SRBs: strict enforcement of the one-child policy can account for 94% of total increase in SRB in 1990, 57% and 54% for the 1991-2000 and 2001-2005 birth cohorts respectively. 7 extra boys per 100 girls for the 1991-2005 birth cohorts. Policy effect passed the robust test, are not confounded by other policy shocks or economic changes. (Li, Yi, & Zhang, 2011)

The explanation is couples are under the constraint of low fertility imposed by the policy, so they tend to perform sex-selection to get a son as earlier as possible. It drives up the SRBs for the first and second orders, contributes most part of overall excess of male births.

Socioeconomic characteristics

Socioeconomic characteristics affect SRB via their various effects on individuals' son preference, access to sex-selective technology and their fertility levels. Attané (2009) analyzes provincial census data, and finds that extreme poverty, family support for elderly and lower fertility together encourage sex-selective behaviors and thus drives up SRB in China. Chuang and Das Gupta 's (2007) study points out the significant effect of development on the decline of son preference in South Korea.

Using the county level data from 2000 census and the 1% intercensal survey of 2005, Guilmoto and Ren (2011) identify an inverted-U relationship between SRB and socioeconomic status, i.e. the poorest and the richest households have lower SRB than those in the middle. Their finding is quite illuminating as previous researches only identify the positive relation between economic well-being and high SRB. For example in India, the relation is linear and increasing (Jha et al., 2011). Some of the motivation for China's strong son preference is economic, but it is not the driving force. We observe high SRBs in both richest provinces (Guangdong and Jiangsu) and poorest provinces (Jiangxi and Guizhou). But less developed regions like Tibet and Xinjiang have SRB close to normal level.

Education

Education is often an explanatory variable for the change in fertility behaviors. It has been identified that improvement of female education plays a robust role in fertility decline (Drèze & Murthi, 2001; Ebenstein, 2010). Also many studies have discussed the effects of education on birth masculinization. But the findings are somehow ambiguous and conflicting.

Gu & Roy (1995) points out a positive link between China's SRB and education in 1990s. Other studies also find that more education is unlikely to ameliorate the shortage of girls (Croll, 2002; Baochang Gu & Li, 1994; Retherford & Roy, 2003). Banister (2004) reports that there was a decline in "missing girls" associated with the decline of illiteracy, but as illiteracy declined further the shortage of girls has become more extreme from late 1970s until now. Guilmoto (2008) finds out Indian women with better education tend to have higher SRB. Explanations for positive education effect on SRB increase include better access to information, more planned reproductive behavior, better financial ability to access medical services and possible upward social mobility through marriage.

A multilevel analysis of SRB in China (with 2000 Census data) points out low education is significantly correlated with high SRB, and therefore increasing education has positive effect of bring SRB back to normal (Guo, 2007). Ebenstein's(2010) comparative study shows that data from India indicates that mothers with higher education appear more likely to have a son than the less educated; while data from China shows that mothers' education is not a significant factor accounting for the male birth. A recent study further shows that the improvement of female education will constantly decrease son preference and China's SRB is about to decline as the socioeconomic improvement (Chen & Hu, 2012). Another recent study shows education also plays a role in reducing SRB in India. The conditional sex ratio for second-order births declines much greater in mothers with 10 or more years of education than those with no education (Jha et al., 2011). There are also parity differentials of education effects. Yang (2006) points out in China mother's education has positive effect on decrease SRB for first order but not significant for second order.

Chung and Das Gupta's (2007) study on the decline of SRB in South Korea is illuminating. They find the drastic decline in son preference is attributable more to social norm spreading through population, rather than the actual increases in proportions of higher education or better socioeconomic status. It has something to do with the diffusion of new ideas. In the first stage, abandonment of son preference happen among the people who are most educated; then it gets spread throughout a society. In this sense, we may not be able to find clear and unidirectional relationship between socioeconomic characteristics (including education) and their underlying and/or manifested son preference, at the individual or household level. But at the societal level, the relationship can be clear. Here, they implicitly point out one critical effect of education on reducing SRB: there should be quite a few people who are highly educated and start to adopt new gender norm. If a new norm can be spread fast through a society, then the effects of education on SRB are actually amplified by passing the new gender norm to the less educated group.

To sum up, education can affect SRB in the following ways:

(1) Higher education (usually alongside with better economic status) provides better information and access to the sex-selective technology, facilitating the realization of attitudinal son preference into reality.
(2) Higher education leads to lower fertility, inducing more sex-selection on earlier orders of births and thus increase the overall SRB.
(3) Prevalence of the sex-selective technology makes education's positive effect on SRB less prominent, where education seems an irrelevant factor.
(4) Higher education makes individuals more open to modern gender norm and abandon son preference, which results in fewer sex-selections and thus reduce the overall SRB.
(5) Higher education can trigger normative change within the society as a whole, rather than just through changes in individuals as their socioeconomic circumstances improved.

Based on the above discussion, the thesis is organized as follows: in chapter 2, we begin with a brief description of the current SRB in China, with information from the latest nationwide census; and then we use the province level and the county level data to explore possible determinants of SRB. In chapter 3, we use individual data from the "China General Social Survey（CGSS）" of 2008 to examine how the sex composition of children is determined at individual and household level. In chapter 4, we will make a brief summary of our findings.

2 SRB Determinants-with 2010 Census Data

2.1 SRB Distribution in China

2.1.1 The 2010 Census Data

The nationwide census is conducted every 10 years in China. The latest one was conducted in 2010. Each person was obliged to report his/her authentic personal and household information to the census investigators. There were two forms of questionnaire used in the census: the short-form and the long-form.

The short-form has 18 questions, 6 of them asking about the household conditions while the rest asking about individual information. Basic demographic information like name, age, gender, level of education, ethnicity/nationality, current residence, registered residence, migration were covered by these questions. While questions about the household include the number of people living within the household, their registered residence, the number of people who was born and die in the past year (between November 1st 2009 and October 31st 2010), the size of residential housing area and the number of rooms in the house/apartment.

The long-form was used to 10% of the total households in China. There were 45 questions. Besides the same information covered by the short-form, the long-form recorded migration, health status, employment status, fertility and other household conditions in finer details.

Information from the short-form and the long-form was organized and sort into two groups of tabulations: the general dataset and the long-form dataset. These data was published on the website of and can be found under the link "Tabulation on the 2010 population census of the P.R.C"(NSB, 2010).

In this thesis, we will use data from two levels: the provincial level and the county level. Each province has its own statistic bureau, which is responsible for summarizing and publishing the 2010 census tabulations in that region. And the tabulations have the same format and structure as those published by NSB. That is tables with the same name contain same variables and these variables are listed in the same order. Therefore, it is easy for us to collect information for the each of our variables, both on the provincial and county level data. We can find 31sets of tabulations of the 2010 census data, one for each province. These tabulations are published on the webpages of the provincial statistic bureaus. And we can find the links from the webpage of National Bureau of Statistics(NSB, 2010).

The provincial data we use include data from all the 31 provinces. These provincial administrative regions include 4 municipalities directly under the central government, 5 autonomous regions and 22 provinces (called “provinces” hereafter).

The county level data contains observations from 1069 counties. County is the lowest administrative level in China. The variables from the tabulations are averages across the population in the county. Some provinces only provide tabulations on the city-level (which is one level higher than county). These data are averages across all the counties, therefore more heterogeneity are lost compared to county data. We thus do not include these data into our sample.

The counties from Jiangxi, Shandong, Shanxi and Tibet have important variables (for our research interest) missing for unknown reasons, therefore are not included in our sample. However, this sample is still very representative for the total since it covers most provinces in China. The observations have considerable diversity in SRB, education and other socioeconomic aspects.

2.1.2 Sex Ratio at Birth

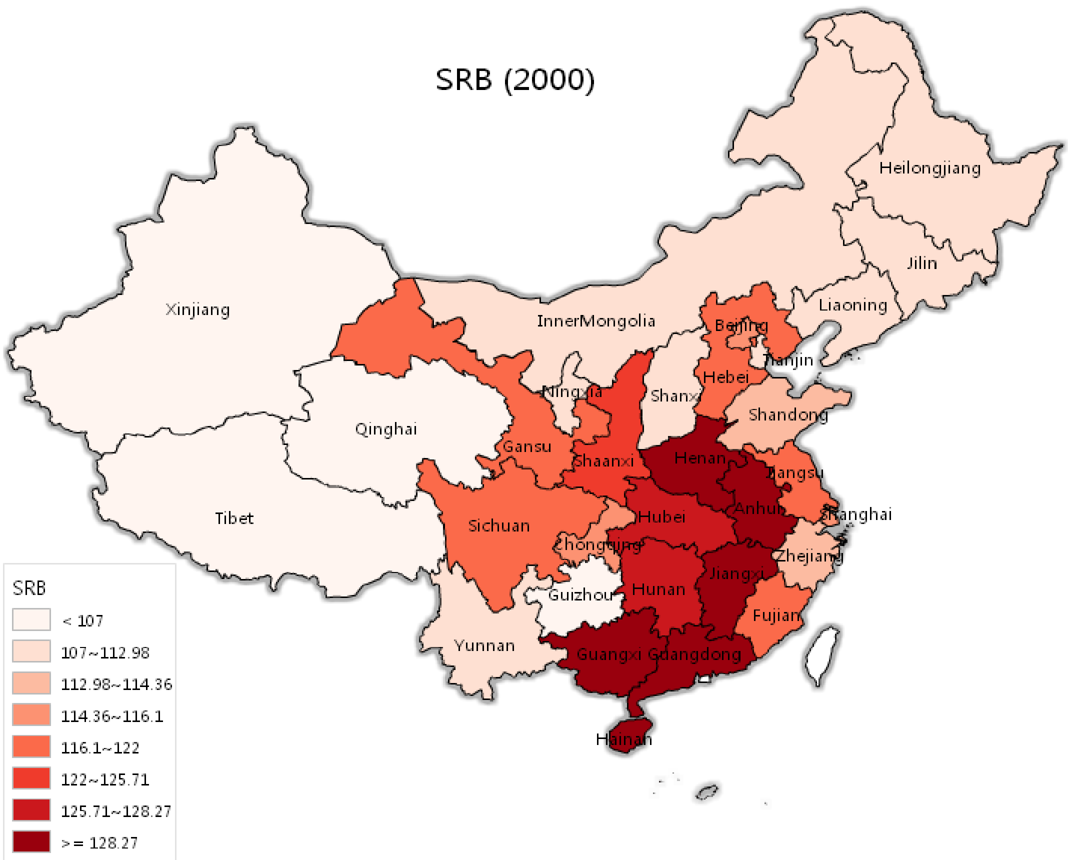
Based on our research interest, the key variable is the Sex Ratio at Birth (SRB). Table 6-1(NSB, 2010) from the long-form dataset presents the levels of SRB directly. Each province also publishes such tabulation under with the same title – Table 6-1(NSB, 2010). We can find the overall SRB and SRB for 5 orders of births of 31 provinces and 1070 counties. Background of these SRBs is the number of boys and girls born between November 1st 2009 and October 31st 2010. The overall SRB is the number of boys born per 100 girls in that province. The SRBs for different orders are ratios of boys to girls among the same order as born to their families. For example, the SRB for first order is the sex ratio computed among the births who were born as the first children in their family.

Table 2.1 The SRB of 31 provinces from 2000 and 2010 census data

	province	srb2010	srb2000	changes
1.	Beijing	112.15	114.58	-2.12
2.	Tianjin	114.59	112.97	1.43
3.	Hebei	118.71	118.46	.211
4.	Shanxi	113.07	112.75	.284
5.	Inner Mongolia	108.87	108.48	.36
6.	Liaoning	112.91	112.17	.66
7.	Jilin	115.67	109.87	5.28
8.	Heilongjiang	115.1	107.52	7.05
9.	Shanghai	111.49	115.51	-3.48
10.	Jiangsu	121.38	120.19	.99
11.	Zhejiang	118.36	113.11	4.64
12.	Anhui	131.07	130.76	.237
13.	Fujian	125.71	120.26	4.53
14.	Jiangxi	128.27	138.01	-7.06
15.	Shandong	124.28	113.49	9.51
16.	Henan	127.64	130.3	-2.04
17.	Hubei	123.94	128.02	-3.19
18.	Hunan	125.78	126.92	-.898
19.	Guangdong	129.49	137.76	-6
20.	Guangxi	122	128.8	-5.28
21.	Hainan	129.43	135.04	-4.15
22.	Chongqing	113.8	115.8	-1.73
23.	Sichuan	112.98	116.37	-2.91
24.	Guizhou	126.2	105.37	19.8
25.	Yunnan	113.61	110.57	2.75
26.	Tibet	100.08	97.43	2.72
27.	Shaanxi	116.1	125.15	-7.23
28.	Gansu	124.79	119.35	4.56
29.	Qinghai	112.69	103.52	8.86
30.	Ningxia	114.36	107.99	5.9
31.	Xinjiang	105.56	106.65	-1.02

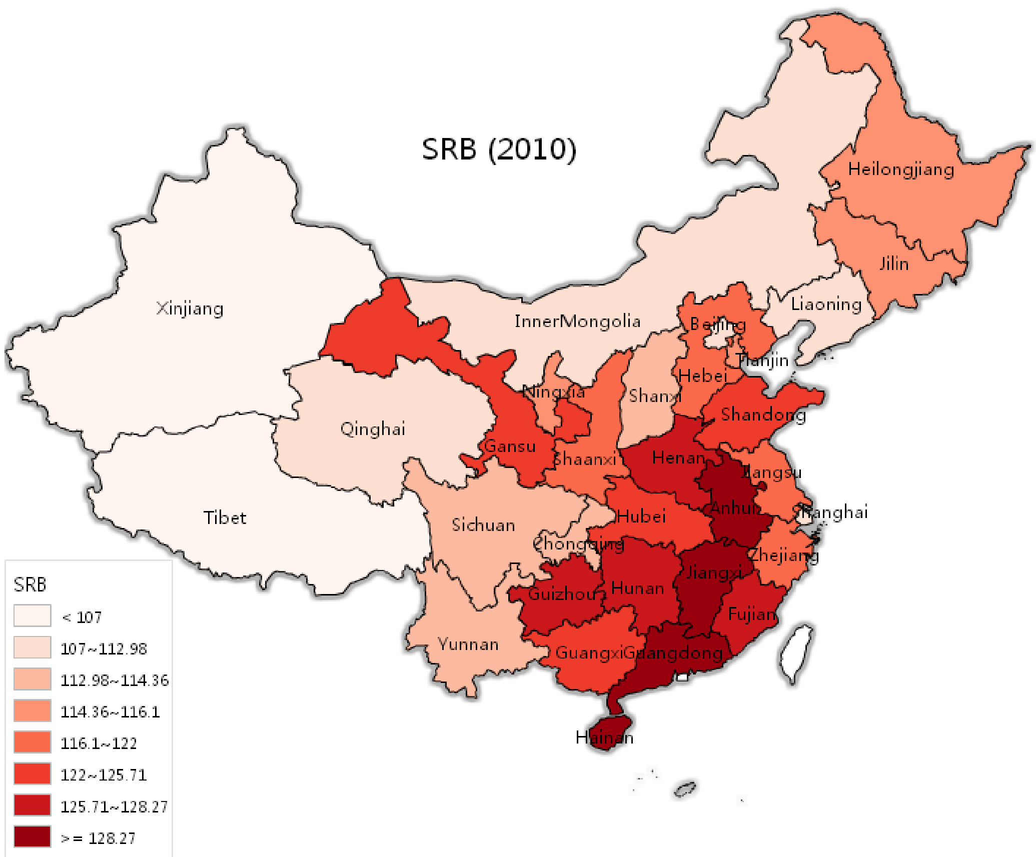
From Table 2.1, we find there are large differences in SRBs across provinces. The highest levels are found in Jiangxi, Guangdong, Hainan, Anhui and Henan where SRBs are above or closely to 130; while the lowest are found in Inner Mongolia, Xinjiang and Tibet, within the normal range for SRB. Over half of the 31 provinces have levels higher than 120, which can be considered as sever birth masculinity. The other provinces have SRBs between 110 and 120, which can be considered as slightly deviations from the normal range.

Figure 2.1 Sex Ratio at Birth across Provinces (2000 Census)



We map the changes in SRB from the 2000 (using the data from the 2000 census (NSB, 2000) to 2010. From Figure 2.1 and Figure 2.2 we can see that the provinces in the Middle and East China have the highest SRBs. And in the West and Northeast, the SRBs are not severely distorted. However, we notice that provinces in the Northeast, Heilongjiang and Jilin, and the province in the Southwest, Guizhou experienced a fast birth masculinization between 2000 and 2010. Especially in Guizhou, the SRB level jumped from 105.37 to 126.2. Some provinces have high SRB in both census years, like Jiangxi, Anhui and Guangdong. In some provinces, the SRB slighted decreased, from a record- high level to a slightly milder level. For example, province Guangxi used to have SRB as high as 128, but the 2010 census data shows the SRB has declined to 122. To sum up, even though, there are some changes in SRBs from 2000 to 2010, the regional differentials continue to be significant.

Figure 2.2 Sex Ratio at Birth across Provinces (2010 Census)



From Table 2.2 we can see that the first order is not far deviated from the normal, even though we observe some differentials across provinces. However, the second order is much above the normal range of SRB. We find Beijing, Anhui, Shandong, Hubei and Guizhou have the highest level for this. We are little shocked to find that Beijing has such a high SRB in the second order of births, since it is the biggest city in China and is supposed to have least preference for son and thus lower SRBs. Most provinces have SRB peaks in the third order of births. SRBs in the later orders are high and contain large variations. They seem shocking but given that most Chinese families are only allowed to have one or two children, these numbers should not account much for the overall SRB. Note that the SRB for the fifth order in Beijing should be an error. We refer to the original table 6-1(NSB, 2010) of the long-form dataset, there were 2 boys and 0 girls born as the fifth child. Therefore, the abnormally large value must be a calculation mistake. It is so far inappropriate to say that the distortion in the first and second orders is most responsible for the overall distortion in SRB. But we do get some hints that there must be some correlations between them and given the weights on the first two orders (births as first and second children account for 93.45% of the total births that the census counted during 2009 and 2010), the correlations are expected to be strong.

Table 2.2 Sex Ratio at Birth and the SRBs for Different orders (2010 Census)

Province	srb	srb_1	srb_2	srb_3	srb_4	srb_5
Beijing	112.15	107.31	137.58	260	900	572157
Tianjin	114.59	111.77	119.77	173.08	325	300
Hebei	118.71	108.98	128.81	184.73	145.34	214.29
Shanxi	113.07	112.51	107.9	165.86	256.41	200
Inner Mongolia	108.87	107.32	109.78	153.33	144.44	600
Liaoning	112.91	110.5	116.64	200.74	153.85	150
Jilin	115.67	116.46	109.21	169.6	92.86	100
Heilongjiang	115.1	115.8	111.85	124.09	87.5	100
Shanghai	111.49	106.35	128.43	201.92	157.14	300
Jiangsu	121.38	114.77	135.01	153.37	136.81	157.14
Zhejiang	118.36	109.7	131.95	192.72	165.31	400
Anhui	131.07	113.01	168.32	245.87	275	342.86
Fujian	125.71	112.7	148.77	240.06	388	100
Jiangxi	128.27	113.89	139.34	167.41	132.03	243.9
Shandong	124.28	113.39	144.76	242.54	207.32	188.89
Henan	127.64	118.05	133.15	175.05	190.55	200
Hubei	123.94	115.01	144.19	171.65	92.86	47.37
Hunan	125.78	119.82	127.85	182.95	151.22	151.72
Guangdong	129.49	122.3	138.72	153.42	142.64	136.29
Guangxi	122	114.61	128.36	140.58	144.11	145.69
Hainan	129.43	117.22	129.88	200.31	204.76	366.67
Chongqing	113.8	111.91	112.28	143	155.88	154.55
Sichuan	112.98	113.67	109.01	123.34	128	112.64
Guizhou	126.2	109.39	146.89	174.33	169.03	154.12
Yunnan	113.61	108.88	116.5	138.93	128.57	137.76
Tibet	100.08	107.75	96.61	98.86	78.26	91.03
Shaanxi	116.1	114.42	116.9	144.97	142.42	400
Gansu	124.79	118.15	131.84	150.34	127.27	225.93
Qinghai	112.69	107.86	115.53	133.33	116.67	115.22
Ningxia	114.36	107.26	121.52	125.29	149.06	165
Xinjiang	105.56	106.22	103.35	108.82	108.49	91.86

2.2 What Determines the SRB Level?

In this part, we will analyze the data collected from tabulations on provinces and counties of the 2010 census(NSB, 2010); to find out what determines SRB differentials across provinces and counties; and how cultural, socioeconomic characteristics affect the levels of SRB, with a special focus on education's effects.

2.2.1 The Variables

Our first dependent variable is SRB. Besides, we also use SRB_1 and SRB_2 as dependent variables, and include the explanatory variables into the models exactly as those in the SRB models.

The reasons for our interests in SRB_1 and SRB_2 include: (1) China's one-child and 1.5-child policy have determined that most Chinese families can only have one or two children, so SRB of the first two orders are relevant with the majority Chinese families; (2) Many studies have pointed out that such policies are responsible for the inordinate SRB (Ebenstein, 2010; Zhu et al., 2009) because the distortion of the first two orders contribute the most part to the excess male births in China; (3) The correlation (see Table 2.3 and Table 2.4) shows the similar result that the first orders are highly correlated with SRB. These inspire us to examine the determinants of the first two orders of SRB. And to find out if the determinations of them are the same.

Table 2.3 Correlation between SRB and its Orders (Provincial Data)

	srb	srb_1	srb_2	srb_3	srb_4
srb	1.0000				
srb_1	0.6888	1.0000			
srb_2	0.8217	0.2735	1.0000		
srb_3	0.4685	0.0468	0.7132	1.0000	
srb_4	0.0204	-0.1858	0.3191	0.6460	1.0000

Table 2.4 Correlation between SRB and its Orders (County Data)

	srb	srb_1	srb_2	srb_3	srb_4
srb	1.0000				
srb_1	0.5334	1.0000			
srb_2	0.7766	0.0520	1.0000		
srb_3	0.3190	-0.0794	0.3493	1.0000	
srb_4	0.1274	-0.0178	0.0505	0.0365	1.0000

Explanatory variables are specified as following:

(1)Total fertility rate (denoted as "tfr" in the models)

The hypothesis is SRB is higher when fertility is low. The reason is couples are more likely to use manipulative technology to get the sex of the child their desired. For example, if a couple decides to have only one child, then with certain degree of son preference they would want to have a son. The values of total fertility rate for each province and each county are listed in Table 6-4 from the long-form dataset(NSB, 2010).

(2)The percentage of population who are minorities ("minority")

The family planning policy in China has less strict control over the fertility behavior of people who are not Han-nationality. And we assume that most non-Han ethnic groups do not have son preference in their culture, or their son preference is not as strong as Han group. So they would not as desperate as Han people to manipulate children's sex. Therefore, we expect that regions with higher composition of minority population will have a lower SRB. The information on ethnic composition of total population is available in Table 1-6 from the general dataset(NSB, 2010). We calculate the variable as the number of non-Han population divided by the total population in that region, times 100 (to get the percentage).

(3)The percentage of multigenerational households ("multgen")

Young couples who live with parents, usually husband's parents, are more likely under the pressure to at least have a son to for the continuity of the family line. Therefore, we assume regions with higher percentage of multigenerational households will have a higher SRB. We find such information from Table 1-11 from the general dataset (NSB, 2010). We just sum the percentages of three-generational and four-generational households, and assign it to our variable. Note that households with five generations living together are very few in each regions, and even smaller in percentage, so we neglect it.

(4)The percentage of urban population ("urban")

We assume that of modernization and urbanization will reduce people's preference for sons, as is happening in South Korea. And the proportion of urban residents is a plausible index to measure the degree of urbanization. Therefore, we expect regions with higher proportion of urban residents will have lower SRBs. Such information can be found in Table 1-5 from the general dataset(NSB, 2010). The percentage of non-rural population is directly list in the table.

(5)Percentage of elderly (above 65) who rely on family support ("fs")

Traditionally, it is sons' duty to take care of parents when they grow older. This is an important reason that sons are more valuable for parents and therefore more preferred. In this sense, we think this variable contains some information on how many families are still adhere to such traditional elderly support form. On the other hand, we also assume that elderly will not rely on the support from their family if they have other source of income, such as pension. In this sense, the variable can also indicates the public elderly supports provided by the local government. If there is a good universal pension system for the elderly, then the traditional family support is not necessary and the economic reason for son preference will be much weakened. Therefore, we assume that regions with lower percentage of the elderly people who rely on family support, will have lower SRBs. Such information is available in Table 8-6 from the long-form dataset(NSB, 2010). The table lists the population of age 65 and above, and the number of them who rely on supports from family members. So we can calculate the percentages to get values for our variable.

(6)The percentage elderly (above 60) require daily care ("care")

Due to the traditional elderly support form, going to nursing home is still not the first option for most Chinese elderly people. Besides, the number of nursing homes and the general quality of nursing staff are still not met by the demand(Chu & Chi, 2008). Therefore, we assume it is still the family members who take care of the unhealthy elderly, and this means an extra burden to the families. And this will account for large amount of family resources that could have been used for raising the next generation. Therefore, young couples will tend to have fewer children, and given some degree of son preference they are more likely to manipulate child's sex. Therefore, we assume regions with larger percentage of unhealthy in the elderly group will have higher SRBs. Information on this can be found in Table 8-1 from the long-form dataset(NSB, 2010).

(7)Percentage of population who have migrated more than 6 months ("migration")

Since large scale of migration will help the new ideas to spread faster through the entire population. In China, famers migrating from rural areas and working in the cities is a phenomenon lasting for about 20 decades. It is a very important demographic phenomenon and consists of a separate category in the 2010 census data. For the variable here, we use Table 1-3from the general dataset(NSB, 2010), in which the percentage of population who have migrated for longer than 6 months is listed. "Have migrated for more than 6 months" means that those people were not living in the place they were registered in, at the point the census was conducted, and had been away from home for more than 6 months, for various reasons. And we assume regions with higher percentage of migration people will have a lower SRB, due to the diffusion of new gender norm against traditional son preference.

(8)Education

The variable we are most interested in is education. We use three variables to describe the education status of each region: Average years of schooling (this is the overall average including both male and female and denoted as "ays"), Female deviation from the overall average years of schooling ("fays") and Male deviation from the overall average years of schooling ("mays"). The first variable describes the general educational level of the population, and the later variables describe the relative education level of female and male groups in the same region. The background information for these variables are found in Table 1-8 from the general dataset(NSB, 2010).

In Table 1-8, there are 7 categories under the item "education": none-education, primary, junior high school, senior high school, college, university and graduate. And it presents the sums of population under each category, with the overall as well as sex-specific sums. Based on these, we can compute for the percentages of people under each category, both the overall and sex-specific ones. In this way, we can get the education composition (in percentages) for the entire population in that region and the separate education compositions for female and male.

We then convert the 7 categories of education into "years of schooling" by the standard years that will it takes to complete that level of education. For "none-education" the number of years of schooling is obviously 0. For "primary", it means 6 years of schooling. To complete "junior high school", it usually takes 9 years. And for "senior high school", it means 12 years. For "college", it usually takes another 3 years, so it means 15 years of schooling in total. It takes usually 4 years to complete a Bachelor degree in China, so the total years of schooling for a Bachelor are 16 years. Students usually study for 3 years in graduate schools to get a Master degree, and another 3 years to get a PhD. Since the original tabulation includes the Masters and PhDs in one group, for simplicity we assume all observations are have Master degrees, which means years of schooling for the highest category "graduate" are 19.

Then we can calculate the "average years of schooling" across the population for each region ("ays"), as well as that among the female population (ays_f) and the male population (ays_m). To describe the relative education levels of female and male. We generate two variables called "female deviation from the overall average years of schooling" ("fays") and "male deviation from the overall average years of schooling" ("mays"). The calculation is simple. We subtract "ays_f" by "ays" to get "fays"; and subtract "ays_m" by "ays" to get "mays". These education variables are candidate explanatory variables, but we need to examine their correlations before we include them into our models.

We can see from Table 2.5, the average female education is lower than the overall level, while the male education is higher than the overall level in all provinces. But we can still find some regional differentials. People in Beijing and Shanghai have the highest level of education. They have 11.35 and 10.17 years of schooling on average. People from Tibet and Guizhou have fewest years of schooling, 4.67 and 6.84 respectively. And in the majority of provinces, people have 7-9 years of schooling on average. As to the sex differentials, we find that women from Beijing, Tianjin, Liaoning, Jilin and Xinjiang have the least deviation from the overall average. That means male and female are more equal in terms of education. But in Anhui, Hainan, Guizhou, Tibet and Gansu, we observe that men are much more privileged than women in terms of education.

(9)Socioeconomic variable

The relationship between the socioeconomic levels and birth masculinity has been studied by many scholars(Attané, 2009; Chung & Das Gupta, 2007; Guilmoto & Ren, 2011) Some of them use one or several variable as proxy for the general socioeconomic status(Attané, 2009; Chung & Das Gupta, 2007). Such way of modeling socioeconomic status is too simple and the variables cannot describe regions' socioeconomic status well enough. Guilmoto and Ren (2011) use a synthesized socioeconomic variable and find out the inverted "U" relationship between SRB and socioeconomic status. But they include education as a part of the socioeconomic status. In this thesis, we want to study the effect of education on SRB, as an independent explanatory variable. We also wonder if education is still a significant variable when socioeconomic status is already incorporated into the model.

We use sector and occupation data from Table 4-4 and Table 4-7 from the long-form dataset(NSB, 2010), to construct our socioeconomic variable. There are 20 categories for sector and 7 categories for occupation. In these tables we can find the number of people for each of these categories. We apply standard Principle Component Analysis(Jolliffe, 2005) to construct a synthetic index for socioeconomic status, so that each region has a socioeconomic score. The Principle Component Analysis is a data reduction method which is used to re-express multivariate data with fewer dimensions. These dimensions should capture the maximum possible information (variation) from the original variables. We start with finding the eigenvectors and eigenvalues of the correlation matrix of the original variables, and then rank them from the highest to the lowest eigenvalues. The eigenvector attached with the highest eigenvalue covers the most information of the correlation matrix. The eigenvector with the next highest eigenvalue contains less information than the first one. Then we know how much they together can explain the variance of the original variables. If they can explain at least 50% of the variance, then we consider using the scores computed based on them is as good as using the original variables. In our provincial data and county level data, we find the first components can explain over 50% of the variance of the original variables. Thus we can compute the socioeconomic score for each region from the 27 original variables, without losing important information about the observations. With this computation, the overall socioeconomic status of a region is denoted by a single score, while higher score means higher socioeconomic status (see Table 2.5).

Table 2.5 Education and Socioeconomic Status of provinces

province	ays	ays_f	ays_m	fays	mays	pca
Beijing	11.46	11.35	11.57	-0.11	0.11	15.27
Tianjin	10.10	9.95	10.23	-0.15	0.13	5.60
Hebei	8.83	8.58	9.07	-0.25	0.24	-2.91
Shanxi	9.17	8.97	9.35	-0.20	0.19	0.01
Inner Mongolia	8.93	8.63	9.21	-0.31	0.28	0.64
Liaoning	9.40	9.21	9.59	-0.19	0.19	1.56
Jilin	9.24	9.06	9.40	-0.17	0.17	-0.85
Heilongjiang	9.11	8.93	9.29	-0.19	0.18	-0.35
Shanghai	10.50	10.17	10.82	-0.33	0.31	10.42
Jiangsu	9.08	8.59	9.56	-0.48	0.48	2.05
Zhejiang	8.57	8.17	8.96	-0.41	0.39	2.53
Anhui	8.08	7.49	8.66	-0.59	0.58	-1.79
Fujian	8.76	8.30	9.20	-0.46	0.44	1.80
Jiangxi	8.52	8.04	8.99	-0.48	0.46	-1.27
Shandong	8.71	8.23	9.19	-0.48	0.47	-2.27
Henan	8.61	8.29	8.94	-0.32	0.32	-3.88
Hubei	8.96	8.50	9.40	-0.46	0.44	-1.16
Hunan	8.86	8.57	9.14	-0.29	0.28	-2.11
Guangdong	9.18	8.83	9.51	-0.35	0.32	1.32
Guangxi	8.40	8.06	8.72	-0.34	0.32	-3.48
Hainan	8.85	8.34	9.30	-0.50	0.45	-0.53
Chongqing	8.49	8.19	8.78	-0.29	0.29	-0.20
Sichuan	8.12	7.78	8.45	-0.34	0.33	-2.49
Guizhou	7.41	6.84	7.95	-0.57	0.54	-3.69
Yunnan	7.54	7.15	7.89	-0.38	0.36	-3.85
Tibet	5.25	4.67	5.79	-0.57	0.54	-4.10
Shaanxi	9.07	8.72	9.39	-0.35	0.33	-0.96
Gansu	7.96	7.41	8.50	-0.55	0.53	-3.80
Qinghai	7.58	7.06	8.06	-0.52	0.48	-0.82
Ningxia	8.44	8.02	8.85	-0.42	0.40	0.19
Xinjiang	8.85	8.74	8.96	-0.11	0.11	-0.88

Table 2.6 Variable Descriptions with Provincial and County level data

tfr	Total fertility rate
minority	Percentage of population with non-han nationality
migration	Percentage of population who migrate longer than 6 months
urban	Percentage of population with urban Hukou
multgen	Multigenerational household
care	Percentage of above 60 who require daily care
fs	Percentage of above 65 population rely on family member support
ays	Average years of schooling
ays2	Squared average years of education
ays_f	Average years of schooling for female
ays_f2	Squared female average years of education
ays_m	Average years of schooling for male
mays	Male deviation from the overall average years of schooling
fays	Female deviation from the overall average years of schooling
pca	Socioeconomic status (PCA)
pca2	Squared Socioeconomic status
srb	Sex ratio at birth
srb_1	Sex ratio at birth_first children
srb_2	Sex ratio at birth_second children

Table 2.7 Descriptive Statistics of Variables (Provincial Data)

Variable	Obs	Mean	Std. Dev.	Min	Max
tfr	31	1186.284	287.0408	706.7	1789.75
minority	31	14.91964	21.3172	.2622276	91.83046
migration	31	19.43677	5.755417	5.31	29.72
urban	31	31.93226	12.56511	14.77	61.89
multgen	31	17.4929	4.714753	9.52	25.22
care	31	3.154652	.7383287	1.797652	5.449395
fs	31	46.82114	12.78878	5.199708	63.06847
ays	31	8.710771	1.040049	5.246712	11.4606
ays_f	31	8.350276	1.137241	4.674943	11.34662
ays_m	31	9.054971	.9577858	5.78741	11.56756
mays	31	.3442005	.1366138	.1066551	.5805206
fays	31	-.3604952	.1420931	-.59059	-.1139822
pca	31	2.34e-09	4.132688	-4.102504	15.26901
srb	31	118.3897	7.707074	100.08	131.07
srb_1	31	112.4832	4.22548	106.22	122.3
srb_2	31	126.1516	15.80838	96.61	168.32
ays2	31	76.92434	17.72131	27.52799	131.3454
ays_f2	31	70.9787	18.6179	21.8551	128.7458
pca2	31	16.52817	44.78749	.000192	233.1428

Table 2.8 Descriptive Statistics of Variables (County Data)

Variable	Obs	Mean	Std. Dev.	Min	Max
tfr	1069	1298.694	394.9167	447.64	4029.81
minority	1069	11.91129	24.30011	0	98.8628
migration	1069	20.27996	9.240628	2.59	95.64
urban	1069	27.05418	20.12846	3.83	93.75
multgen	1069	20.06116	8.265498	.83	84.31
care	1069	3.034576	1.798435	0	29.20792
fs	1069	52.51631	16.6671	0	88.93113
ays	1067	8.382933	1.269001	3.84	12.84
ays_f	1068	7.931657	1.361677	3.04	12.56
ays_m	1068	8.815974	1.197073	4.15	13.12
mays	1067	.4313308	.1679356	.01	1.13
fays	1067	-.4533271	.171137	-1.15	-.02
pca	1066	-4.17e-10	3.619252	-5.101148	15.81012
srb	1068	121.6089	23.90153	0	300
srb_1	1068	115.3689	31.00124	0	800
srb_2	1064	139.3649	56.47212	20	700
ays2	1067	71.88243	22.08099	14.7456	164.8656
ays_f2	1068	64.76361	22.52859	9.2416	157.7536
pca2	1066	13.08669	26.33133	.0000786	249.9599

Table 2.9 Correlation Matrix of dependent and independent variables (provincial data)

(obs=31)

	tfr	minority	migrat~n	urban	multgen	care	fs	ays	ays_f	ays_m	mays	fays	pca
tfr	1.0000												
minority	0.3470	1.0000											
migration	0.0224	-0.3411	1.0000										
urban	-0.6712	-0.2519	0.2305	1.0000									
multgen	0.4007	0.0496	-0.4237	-0.6940	1.0000								
care	-0.0725	0.5099	-0.2374	0.0565	0.0258	1.0000							
fs	0.5212	0.1647	-0.2774	-0.9139	0.7190	-0.1689	1.0000						
ays	-0.4943	-0.6589	0.3439	0.7684	-0.5357	-0.2631	-0.7147	1.0000					
ays_f	-0.5121	-0.6269	0.3069	0.7791	-0.5507	-0.2433	-0.7260	0.9955	1.0000				
ays_m	-0.4699	-0.6913	0.3819	0.7467	-0.5117	-0.2844	-0.6934	0.9940	0.9792	1.0000			
mays	0.4688	0.1695	0.0597	-0.6151	0.4907	0.0092	0.5793	-0.6440	-0.7136	-0.5567	1.0000		
fays	-0.4808	-0.1951	-0.0609	0.6114	-0.4869	-0.0215	-0.5798	0.6477	0.7172	0.5610	-0.9974	1.0000	
pca	-0.6179	-0.3131	0.4123	0.8003	-0.6177	0.1047	-0.7597	0.7780	0.7640	0.7838	-0.4279	0.4200	1.0000

Table 2.10 Correlation Matrix of dependent and independent variables (county data)

(obs=1064)

	tfr	minority	migrat~n	urban	multgen	care	fs	ays	ays_f	ays_m	mays	fays	pca
tfr	1.0000												
minority	0.2393	1.0000											
migration	0.1024	-0.2368	1.0000										
urban	-0.4364	-0.2166	0.1192	1.0000									
multgen	0.2391	0.2071	-0.4707	-0.4580	1.0000								
care	0.2420	0.4141	-0.1285	-0.2713	0.1352	1.0000							
fs	0.3988	0.1943	-0.0896	-0.7867	0.4857	0.1525	1.0000						
ays	-0.3589	-0.4229	0.1529	0.5055	-0.2894	-0.4258	-0.4615	1.0000					
ays_f	-0.3642	-0.4071	0.1460	0.4951	-0.2917	-0.4164	-0.4610	0.9940	1.0000				
ays_m	-0.3487	-0.4397	0.1601	0.5101	-0.2808	-0.4328	-0.4545	0.9924	0.9734	1.0000			
mays	0.2275	0.0634	-0.0148	-0.1873	0.1865	0.1348	0.2508	-0.4860	-0.5760	-0.3751	1.0000		
fays	-0.2352	-0.1028	0.0255	0.1892	-0.1726	-0.1546	-0.2442	0.4916	0.5841	0.3842	-0.9770	1.0000	
pca	-0.5023	-0.2692	0.1958	0.8698	-0.4952	-0.3210	-0.7270	0.5495	0.5398	0.5537	-0.2089	0.2180	1.0000

2.2.2 Multiple Regression Model

We first regress SRB on the explanatory variables we specified above. Then we regress srb_1 and srb_2, using exactly the same explanatory variables on the right hand side (RHS) of the equations. The reason we model in this way is we assume all the RHS variable have some explaining power for both SRB and its first two orders. And we expect these variables have different effects on SRB and its first two orders. Models specified in this way will make comparison feasible and easy to interpret. The multiple regression models are specified as following:

$$\boldsymbol{SRB_{mhi}} = \boldsymbol{\beta}_{mh}\boldsymbol{X'_h} + \boldsymbol{e_{mhi}}$$

Where, i=1,…n (n denotes the number of observations)

m=0, 1, 2. For m=0 the dependent variable is SRB; m =1 the dependent variable is SRR_1; m=2 the dependent variable is SRB_2

$\boldsymbol{X_h}$ are the explanatory variable vectors, with h=1, 2, 3, 4, 5 denoting there are five vectors of explanatory variables for each dependent variable.

To lower the risk of multicollinearity, we exclude some variables from the models. For example, with provincial data (see Table 2.9), "urban" and "fs" are highly correlated with education variables and socioeconomic score ("pca"), therefore are excluded from the models. We observe that correlation between "pca" and education variables are also high, therefore we should avoid modeling them into the same equation. $\boldsymbol{X_h}$ for models with provincial data are specified as following:

X_1= (1 migration care multgen minority tfr)
X_2= (1 migration care multgen minority tfr pca pca2)
X_3= (1 migration care multgen minority tfr ays ays2)
X_4= (1 migration care multgen minority tfr ays ays2 ays_f)
X_5= (1 migration care multgen minority tfr ays_f ays_f2)

With county data (see Table 2.10), we also find the correlations between “fs”, “urban” and “pca” are high, so it is better to leave them out of our models. But we do not observe high correlations between “pca” and education variables, therefore we can include them in the same models. $\boldsymbol{X_h}$ for models with county data are specified as following:

X_1= (1 migration care multgen minority tfr)

X_2= (1 migration care multgen minority tfr pca pca2)

X_3= (1 migration care multgen minority tfr pca ays)

X_4= (1 migration care multgen minority tfr pca ays ays2)

X_5= (1 migration care multgen minority tfr pca ays_f ays_f2)

$\boldsymbol{\beta}_{\boldsymbol{mh}}$ are the coefficient vectors, whose elements are corresponding to the h-th explanatory variables/equation for the m-th dependent variable. The number of elements in the vector depends on which explanatory variable vector is using in the regression.

$\boldsymbol{e_{mhi}}$ are the disturbances. It denote the disturbance in the h-th equation for the m-th dependent variable.

According to Hill, Griffiths and Lim(2008), our multiple regression models have the following assumptions:

(1) $E(\boldsymbol{e_{mhi}}|\boldsymbol{X}) = 0$

(2) $Cov(\boldsymbol{e_{mhi}}, \boldsymbol{e_{mhj}}) = 0 \; for \; i \neq j$

(3) $Var(\boldsymbol{e_{mhi}}|\boldsymbol{X}) = \boldsymbol{\sigma^2_{mh}}$

(4) $No\ correlation\ among\ disturbances\ across\ equations.$

(5) $Explanatory\ variables\ are\ not\ exactly\ collinearly\ related.$

For regressions with provincial data, the number of observations is 31. But for those with the county level data, the number of observations can vary slightly from equation to equation. The reason is for some variable that are included in the models, we do not have data for it in some counties. Or there are some statistic errors in the original tables from the dataset.

2.2.3 Findings and Discussions

<u>Results with the provincial data</u>

Table 2.11 Multiple Regression on SRB with Provincial Data

	(1) srb	(2) srb	(3) srb	(4) srb	(5) srb
migration	0.0487	0.00495	-0.0932	-0.189	-0.103
	(0.170)	(0.182)	(0.145)	(0.136)	(0.140)
care	-2.722*	-3.908*	-5.544***	-4.926***	-5.644***
	(1.269)	(1.548)	(1.298)	(1.197)	(1.255)
multgen	0.338	0.345	-0.00512	-0.0175	-0.0295
	(0.214)	(0.243)	(0.228)	(0.206)	(0.218)
minority	-0.208***	-0.194**	-0.355***	-0.290***	-0.361***
	(0.0520)	(0.0535)	(0.0656)	(0.0648)	(0.0600)
tfr	0.0138***	0.0141**	0.0216***	0.0185***	0.0213***
	(0.00358)	(0.00421)	(0.00359)	(0.00347)	(0.00339)
pca		-0.183			
		(0.526)			
pca2		0.0412			
		(0.0387)			
ays			-32.46***	-5.238	
			(8.542)	(13.34)	
ays2			1.791**	1.326**	
			(0.477)	(0.469)	
ays_f				-17.12*	-28.10***
				(6.850)	(6.771)
ays_f2					1.598***
					(0.396)
_cons	106.8***	110.0***	262.5***	206.9***	240.1***
	(6.881)	(8.193)	(41.77)	(43.76)	(32.36)
N	31	31	31	31	31
adj. R-sq	0.711	0.709	0.807	0.843	0.822

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

We find that socioeconomic status (pca) and its quadratic form are not significant while other variables are in the model. In factual P-values for variable "pca" and "pca2" (squared pca) are 0.73 and 0.3 (not shown in Table 2.9), far below any significance level. Therefor with provincial data, we do not found the inverted "U" socioeconomic effect suggested by Guilmoto and Ren's studies (2011). But we find that education's effect is significant. In the third model, we include both "ays" and its quadratic form "ays2" and find significant negative coefficients, indicating a nonlinear effect of education on SRB. The estimated partial effect of education when "ays"=9 can be calculated as: $\frac{\partial SRB}{\partial ays}|_{ays=9} = -32.46 + 1.79 \times 9 = -16.35$. That mean if the average years of schooling increase from 8.5 to 9.5, then SRB is expected to decline by16.35 while other variables are held constant. Such negative partial effect of education on SRB is expected to be effective until $ays = 19$, i.e. everyone has a Master degree. It also implies such negative effect is more prominent when education increases from a lower level. After we add female education "ays_f", the nonlinear effect is still significant and the coefficient to "ays_f" is also significant. It means educational level of female has a significant negative effect on SRB. One year increase in female relative education will lead to a decline in SRB by 17.12. Then we include both female education and its quadratic form into the model. And we observe a significant negative nonlinear relation between female education and SRB. Such nonlinear relation is very similar to that between the overall education and SRB, only with smaller magnitude. Besides, the results show that SRB is increasing in fertility. This is the opposite as we expected. We assumed lower fertility would increase SRB. Large percentage of minority in the population is on average associated with a lower SRB. The effect of "care" is also against our hypothesis. We expected that increase in elderly people who need daily care will cut down family sources that would have been used to raise another child; and a lower fertility should be associated with a higher SRB. But our data shows percentage of elderly who requires daily care is expected to lower SRB. And as we expected, SRB is decreasing in higher minority composition in the population.

Table 2.12 Multiple Regression on SRB_1 with Provincial Data

	(1) srb_1	(2) srb_1	(3) srb_1	(4) srb_1	(5) srb_1
migration	-0.0869	-0.0474	-0.110	-0.0864	-0.108
	(0.123)	(0.133)	(0.132)	(0.139)	(0.132)
care	-2.457*	-2.668*	-3.027*	-3.181*	-3.087*
	(0.919)	(1.136)	(1.183)	(1.226)	(1.188)
multgen	0.396*	0.305	0.356	0.359	0.366
	(0.155)	(0.178)	(0.208)	(0.211)	(0.206)

minority	-0.0588	-0.0630	-0.0798	-0.0962	-0.0780
	(0.0376)	(0.0392)	(0.0597)	(0.0663)	(0.0568)
tfr	-0.000458	-0.00220	0.00106	0.00182	0.00111
	(0.00260)	(0.00309)	(0.00327)	(0.00355)	(0.00321)
pca		-0.425			
		(0.386)			
pca2		0.0227			
		(0.0284)			
ays			-6.045	-12.84	
			(7.786)	(13.66)	
ays2			0.346	0.462	
			(0.435)	(0.480)	
ays_f				4.276	-5.356
				(7.015)	(6.406)
ays_f2					0.323
					(0.375)
_cons	116.4***	119.7***	143.9***	157.8**	139.5***
	(4.983)	(6.009)	(38.07)	(44.82)	(30.62)
N	31	31	31	31	31
adj. R-sq	0.495	0.479	0.466	0.451	0.469

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

For SRB of the first order, neither education variables nor socioeconomic variables seem to be significant. The only significant coefficient is found with "care". It has negative effect on SRB_1, similar to the effect on SRB but at a lower significance level and smaller magnitude. Such result indicates that our model is not very realistic in modeling the determination of SRB_1.

Table 2.13 Multiple Regression on SRB_2 with Provincial Data

	(1) srb_2	(2) srb_2	(3) srb_2	(4) srb_2	(5) srb_2
migration	0.480	0.181	0.00345	-0.355	-0.0201
	(0.516)	(0.520)	(0.397)	(0.324)	(0.383)
care	-0.0669	-4.983	-9.921*	-7.602*	-10.09**
	(3.843)	(4.427)	(3.560)	(2.847)	(3.439)
multgen	0.0499	0.311	-1.044	-1.090*	-1.126
	(0.649)	(0.695)	(0.626)	(0.490)	(0.596)
minority	-0.424*	-0.351*	-0.908***	-0.661***	-0.925***
	(0.157)	(0.153)	(0.180)	(0.154)	(0.164)
tfr	0.0264*	0.0321*	0.0532***	0.0416***	0.0518***
	(0.0109)	(0.0120)	(0.00984)	(0.00824)	(0.00930)
pca		0.214			
		(1.504)			

pca2		0.133			
		(0.111)			
ays			-111.5***	-9.351	
			(23.43)	(31.73)	
ays2			6.199***	4.453***	
			(1.309)	(1.116)	
ays_f				-64.27***	-95.17***
				(16.29)	(18.55)
ays_f2					5.444***
					(1.086)
_cons	91.12***	97.94***	620.8***	412.2***	538.7***
	(20.84)	(23.42)	(114.6)	(104.1)	(88.65)
N	31	31	31	31	31
adj. R-sq	0.369	0.435	0.655	0.789	0.682

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

Socioeconomic scores still do not affect the sex ratio of birth. But we observe a very strong effect of education on reducing SRB_2, much stronger than it has on SRB. The estimated partial effect of education when "ays"=9 can be calculated as: $\frac{\partial SRB\_2}{\partial ays}|_{ays=9} = -111.5 + 6.2 \times 9 = -55.7$. It means if the average years of schooling increase from 8.5 to 9.5, then SRB for the second order is expected to decline by 55.7 while other variables are held constant. This is an extraordinarily big effect. And it is effective until ays = 18. The nonlinear pattern is not significant. Female education also seems to have such effect, only with smaller magnitude. The reason that education has such effect on reducing the male births among the second children, is higher education reduces people's son preference thus sex-selection on children is less prevailed.

Besides, we find higher total fertility rate is associated with a higher SRB_2. Provinces with larger minority people tend to have lower SRB_2. And provinces with large care demand from the elderly, are expected to have lower SRB_2s.

<u>Results for the county-level data</u>

Table 2.14 Multiple Regression on SRB with County-Level Data

	(1) srb	(2) srb	(3) srb	(4) srb	(5) srb
migration	0.182	0.206*	0.205*	0.203*	0.201*

	(0.0974)	(0.0973)	(0.0972)	(0.0972)	(0.0971)
care	-0.743	-1.101*	-1.251**	-1.371**	-1.409**
	(0.448)	(0.462)	(0.465)	(0.487)	(0.484)
multgen	0.0422	-0.116	-0.108	-0.110	-0.115
	(0.105)	(0.116)	(0.112)	(0.112)	(0.112)
minority	-0.127***	-0.131***	-0.144***	-0.157***	-0.159***
	(0.0341)	(0.0343)	(0.0352)	(0.0381)	(0.0375)
tfr	0.00546**	0.00193	0.00178	0.00165	0.00150
	(0.00205)	(0.00225)	(0.00224)	(0.00224)	(0.00224)
pca		-1.091**	-0.786**	-0.830**	-0.818**
		(0.403)	(0.284)	(0.289)	(0.287)
pca2		0.0168			
		(0.0441)			
ays			-1.326	-5.855	
			(0.741)	(5.442)	
ays2				0.256	
				(0.304)	
ays_f					-5.636
					(4.567)
ays_f2					0.247
					(0.269)
_cons	113.7***	122.0***	134.0***	154.3***	152.8***
	(3.717)	(4.351)	(8.077)	(25.53)	(20.70)
N	1068	1065	1063	1063	1064
adj. R-sq	0.032	0.042	0.044	0.044	0.046

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

Note that even though our county sample size is 1069, but due to missing values for some variables, we lost a few but not many observations (at most 6 are lost in above estimations). The reason could be statistic mistake or some extreme situations. For example, a very small county (an island) in Hainan province did not have births during the observation time period of the 2010 census. Therefore, its SRBs are missing.

Results from the county-level data indicate a significant socioeconomic effect on SRB. Counties with higher socioeconomic scores are associated with lower SRB. For one unit increase on socioeconomic score, the county is expected to have a decline on SRB by about 1. We notice that even though the effect is statistically significant, it is not a strong effect. Referring to Table 2.3, we find that except for some big cities like Beijing and Shanghai (who “pca” scores are 15.27 and 10.42) most provinces have scores between 1 and 4. Since county is the lowest administrative level in China, and most counties are not as developed as cities, we can expect their "pca" scores would group in the range 1 and 4. It means for one unit increase on "pca" score, the sector and occupation structure should have a substantial transformation and upgrade. Therefore, we admit economic development can gradually reduce the SRB in that region, but it may take much efforts and possibly much time to attain that effect. Hence, it is not efficient to rely on economic development for its negative effect on SRB.

Education variables lost significance with our county data. Total fertility rate (tfr) is no longer significant but "minority" is still significant. And "care" also stands out as a significant factor: counties with higher proportion elderly who require daily care are expected to have lower SRBs.

In the determination of SRB_1 (Table2.13) we find neither socioeconomic status nor education have significant effects. "care" has opposite sign as it does in the SRB model. As we assumed, large demand of daily care from the elderly increases family burden, and will decrease the number of births and thus end up with higher SRB given some degree of son preference.

Table 2.15 Multiple Regression on SRB_1 with County-Level Data

	(1) srb_1	(2) srb_1	(3) srb_1	(4) srb_1	(5) srb_1
migration	0.0930 (0.128)	0.0996 (0.128)	0.0957 (0.128)	0.100 (0.128)	0.0992 (0.128)
care	1.527** (0.587)	1.695** (0.610)	1.458* (0.615)	1.816** (0.642)	1.742** (0.639)
multgen	0.0753 (0.137)	0.147 (0.153)	0.105 (0.148)	0.108 (0.148)	0.109 (0.148)
minority	-0.144** (0.0447)	-0.134** (0.0453)	-0.150** (0.0465)	-0.114* (0.0503)	-0.122* (0.0496)
tfr	0.00151 (0.00268)	0.00241 (0.00296)	0.00192 (0.00295)	0.00233 (0.00296)	0.00222 (0.00296)
pca		0.589 (0.532)	0.298 (0.375)	0.430 (0.381)	0.448 (0.380)
pca2		-0.0612 (0.0582)			
ays			-0.891 (0.980)	12.62 (7.182)	
ays2				-0.762 (0.401)	

ays_f					9.204
					(6.033)
ays_f2					-0.610
					(0.355)
_cons	107.1***	104.6***	113.8***	53.09	71.02**
	(4.877)	(5.741)	(10.68)	(33.69)	(27.34)
N	1068	1065	1063	1063	1064
adj. R-sq	0.009	0.008	0.008	0.010	0.010

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

In the determination of SRB_2 (Table 2.14), we find socioeconomic status is an irrelevant factor. And the nonlinear effect of education is significant. It means counties with better education tend to have a much lower SRB among the second children. We can calculate the partial effect of education: $\frac{\partial SRB\_2}{\partial ays}|_{ays=9} = -33.64 + 1.74 \times 9 = -17.98$. Even though the effect is much smaller than that suggested with provincial data, it is discernable and relatively large. Similar effect is found with female education. Therefore, according to our data, improvement in people's education will substantially reduce SRB for the second order.

Besides, migration seems to be an important factor. Large population mobility is associated with higher SRB on the second births. The effect was the opposite as we supposed. We assumed large population mobility will help the new gender norm to spread across population and thus reduce son preference and result in lower SRB. But here, works in different ways. Minority is still significant in decrease SRB for the second order. The sign of variable "care" here is negative again. It means counties with higher proportion elderly who require daily care tend to have lower SRBs among all the second births, and such effect is robust even though other variables are considered.

Table 2.16 Multiple Regression on SRB_2 with County-Level Data

	(1)	(2)	(3)	(4)	(5)
	srb_2	srb_2	srb_2	srb_2	srb_2
migration	0.480*	0.503*	0.497*	0.488*	0.482*
	(0.234)	(0.234)	(0.234)	(0.234)	(0.234)
care	-2.191*	-2.602*	-2.994**	-3.839**	-3.828**
	(1.073)	(1.113)	(1.117)	(1.170)	(1.163)
multgen	-0.133	-0.350	-0.358	-0.367	-0.384
	(0.250)	(0.277)	(0.268)	(0.268)	(0.268)

minority	-0.169*	-0.168*	-0.203*	-0.286**	-0.281**
	(0.0812)	(0.0820)	(0.0841)	(0.0909)	(0.0895)
tfr	0.00275	-0.00222	-0.00283	-0.00363	-0.00394
	(0.00491)	(0.00540)	(0.00538)	(0.00538)	(0.00538)
pca		-1.354	-0.962	-1.273	-1.212
		(0.963)	(0.680)	(0.690)	(0.687)
pca2		-0.00375			
		(0.105)			
ays			-2.912	-33.64**	
			(1.777)	(13.00)	
ays2				1.736*	
				(0.728)	
ays_f					-27.23*
					(10.92)
ays_f2					1.421*
					(0.644)
_cons	137.5***	149.0***	176.1***	313.9***	281.4***
	(8.845)	(10.37)	(19.34)	(60.89)	(49.37)
N	1064	1061	1059	1059	1060
adj. R-sq	0.021	0.023	0.025	0.030	0.030

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

2.2.4 Discussions

With provincial data, we find that education has significant and strong effect on reducing SRB. It is a nonlinear effect which will be effective until average education level reaches to a very high level. It also seems to have stronger effect when education improves from a lower level. We find that education has extraordinary negative on reducing the excess male births among the second children. One possible reason is higher education will reduce the son preference among people, and thus reduce the sex-selective abortion on the second births. Such effect is very meaningful since most distortion in SRB is from the abnormally high levels among the second births. If the education level can be improvement in the “1.5-child” regions, substantial reduce in second order SRB seems very promising.

However with county data, we do not discern a significant effect of education, except for on the second order. It means if the average years of schooling increase from 8.5 to 9.5, then SRB for the second order is expected to decline by 17.98 while other variables are held constant. It seems smaller compared by the result with provincial data. But we find this magnitude more realistic. And this finding testify what we found with provincial data, therefore, we consider education's negative effect on SRB is reliable.

With both data, we find that female education works similarly as overall education on SRB and SRB_2. With provincial data, female education has a significant effect on reducing SRB. This effect is prominent for SRB and SRB_2. When women are better educated as men, they have better chance to participate in formal labor market and be economically independent. In a region where women show stronger competiveness and independency, people will show more respect to women and also daughters. The other reason is the places where women are better educated coincide with those have lower son preference and thus lower SRBs. In either way, the connection between female education and SRB is very strong. Thus, improvement on female education is going to reduce both the overall SRB and the SRB for second order.

Socioeconomic status of a region does not have much effect on its SRB level, suggested by provincial data. But it tends to show significant negative effect on SRB with county data. The negative is small even though it shows significance. And we think it is not efficient to rely on economic development for its negative effect on SRB. The main reason is it takes much efforts and possibly much time to attain a substantial transformation and upgrade on the sector and occupation structure. Therefore, we admit economic development can gradually reduce the SRB in that region, but question its efficiency.

We find that migration/population mobility is an important factor for high SRB_2. Our earlier assumption was population mobility can help new social norms spread across population, reducing the overall son preference and therefore leads to a lower level of SRB. But with our data, it seems more population mobility is associated with higher SRB_2. One possible reason is largest composition of population migration finds among the rural residents migrating to big cities for job. They have strong son preference than urban people. And most of these people are under the "1.5-child" policy, people are most likely to apply sex-selective abortion to ensure a boy as the second child. And currently living in the cities gives them better access to sex-selection services.

Total fertility rate has a positive effect on SRB with provincial data but is not a significant factor suggested by county data. It implies that lower fertility is not the main reason for the rise in SRB.

And both provincial data and county level data suggest substantial amount of SRB differentials can be explained by the composition of minority in the population. It is consistent with our expectation as generally speaking ethnic groups do not have strong preference for sons. And they are not constrained by the "one-child" or "1.5-child" policy. The burden from supporting the elderly family members also seems to be relevant for SRB differentials.

Finally, such regional analyses have limitations. The main problem here is that we use the aggregate data to make inferences about relationships at the individual level. And the results we get from such analyses usually provide biased estimates(Steel, Tranmer, & Holt, 2006). Therefore, it seems analysis with individual data would be more reliable.

2.2.5 Validity of the Estimation

Multicollinearity

We have included several explanatory variables into the models. There comes with a risk that some of them may have exact or high linear relationships. If this is the exact collinearity exists, then the least squares estimator is not defined. But it is not a usual case. More often, we may two or more explanatory variables have high but not exact linear dependencies. Even though these circumstances do not violate the least squares assumptions, and the least squares estimator is still the best linear unbiased estimator, the quality of the estimation can be poor. Estimates can be very sensitive for the inclusion or exclusion of some observations.

Our models do not seem to have such problem. First, Table 2.5 and Table 2.6 indicate that our explanatory variables do not have strong correlations in general. For those with high correlation, we chose very carefully. For example, we did not include "pca" and education variables into model together since they are highly correlated with provincial data (0.78). Second, in each of the regression results tables, we list the standard error for each coefficient and we do not observe any widely inflated standard errors. After each of the above regressions, we check for the problem of multicollinearity. Third, we applied multicollinearity test after the each regression to using the **vif** command in STATA. **vif** is short for "variance inflation factor". If a variable has VIF values larger than 10, then it may have a linear correlation with other independent variables and we should take a second thought whether or to include it into the model. Table 2.17 is an example. None of the variables has VIF value higher than 10. Multicollinearity is not a big problem for our models. Our estimations from the multiple regression models are in general stable.

Table 2.17 Test for Multicollinearity

Variable	VIF	1/VIF
pca	2.06	0.485664
ays	1.72	0.581472
multgen	1.66	0.601279
tfr	1.52	0.655916
migration	1.48	0.677966
minority	1.42	0.706272
care	1.35	0.739077
Mean VIF	1.60	

<u>Errors in the data</u>

The reported number of births is not without query. Goodkind (2011) points out that approximately 19% of children at 0-4 ages were not reported by their parents during the 2000 census. Among them there was excess underreporting of little girls, which rendered the recorded SRB considerably higher than the actual level. China's birth planning system contributes to a rise in excess underreporting of daughters relative to sons. This is because violating the family planning policy means fines on parents. Therefore, they have motive to avoid reporting the new born child, especially when it is a girl. Many studies have analyzed such problem (Zhai & Yang, 2009) Some study points out that the main cause of increase in reported SRB in early 1990s, was underreporting of female births(Zeng et al., 1993).

In this thesis, we do not test for such possibility because we believe that the census tabulations have sufficiently good quality. Besides, the levels of SRB are used as the dependent variable in our models. Some measurement errors in the dependent variable will not cause the least squares estimators to be inconsistent or biased.

3 Education and the degree of son preference-with CGSS data

3.1 The CGSS Data

The data we use in this part is from the “China General Social Survey（CGSS）” of 2008. The survey was conducted under a research project called "China General Social Survey", which was sponsored by the China Social Science Foundation. This research project was carried out by Department of Sociology, Renmin University of China & Social Science Division, Hong Kong Science and Technology University, and directed by Dr. Li Lulu & Dr. Bian Yanjie (Yanjie & Lulu, 2008). Before the one in 2008, they conducted four CGSS surveys in 2003, 2004, 2005 and 2006 respectively. In each of these surveys, samples with sizes up to ten thousand were randomly collected from urban and rural areas in China. These surveys cover many aspects of the informants’ lives, including basic demographic information, education and work experience and even their opinions over some “Social hotspots”. Data from 2003 to 2008 are all available now, which can be obtained for registered users (CGSS, 2008).

The CGSS 2008 data we use in this thesis is collected from the most recent survey. There are 6000 observations in the sample. The sample is representative because it covers every province, having observations from both urban and rural areas. We have 3982 observations from urban areas and 2018 observations from rural areas. The survey had respondents with great variety in age, level of education, occupation and other personal and household characteristics.

The CGSS 2008 survey contains two questionnaires. Questionnaire A was designed to document respondents’ basic information, like residence area, sex, age, education, employment status and job positions. By the end, there were a serial questions asking about respondents’ current and previous experiences in education, work and other social and economic activities. Questionnaire B was designed to get information about respondents’ attitudes over globalization and some controversial issues such as social inequality and injustice. With these details, we hope to find out which characteristics, experience and attitudes of the respondents may define their family building-up behaviors.

Besides, the survey was not aimed to investigate respondents' fertility behaviors and the conductor was not the government. So we suppose the respondents had less motivation to underreport the number of their children or cover up the sex composition of their children. We consider the information on birth behavior is less distorted than census data, and shall be accountable for our research.

3.2 Basic information

On average, the urban residents tend to have smaller family size than the rural residents. Since the one-child policy is only implemented in the cities, so the overall fertility is of course lower in the cities. On the other hand, the modern life style is usually associated with lower fertility as is observed in Japan and many developed European countries. So the spontaneous desire for smaller family is also a very important reason for lower fertility. Even though analyses on macro data suggest that lower fertility is not the main reason for the rise in SRB, we wonder if it is a critical factor at individual level. The frequencies of number of children from our sample are summarized in Figure 3.1.

From the survey data we find among families with only one child, there are about 139 boys to 100 girls. For families with two children, the ratio is much more balanced, about 117 boys to 100 girls. When family size gets to three, the ratio (104.52) falls in the normal interval of sex ratio at birth. Our sample shows 235 families have four children. Among these families, there are 448 boys and 492 girls, which lead to a fairly low boy to girl ratio, 91.1. Associated with family size from five to seven, we observe a slightly increase in the sex ratio of children, but it never reaches a level as high as for the two smallest family sizes. In the families with eight or nine children, there are much fewer boys compared to girls, which draw the sex ratio to an abnormally low level, 71 and 67 respectively (See Figure 3,2).

Figure 3.1 Observations on Number of Children (CGSS)

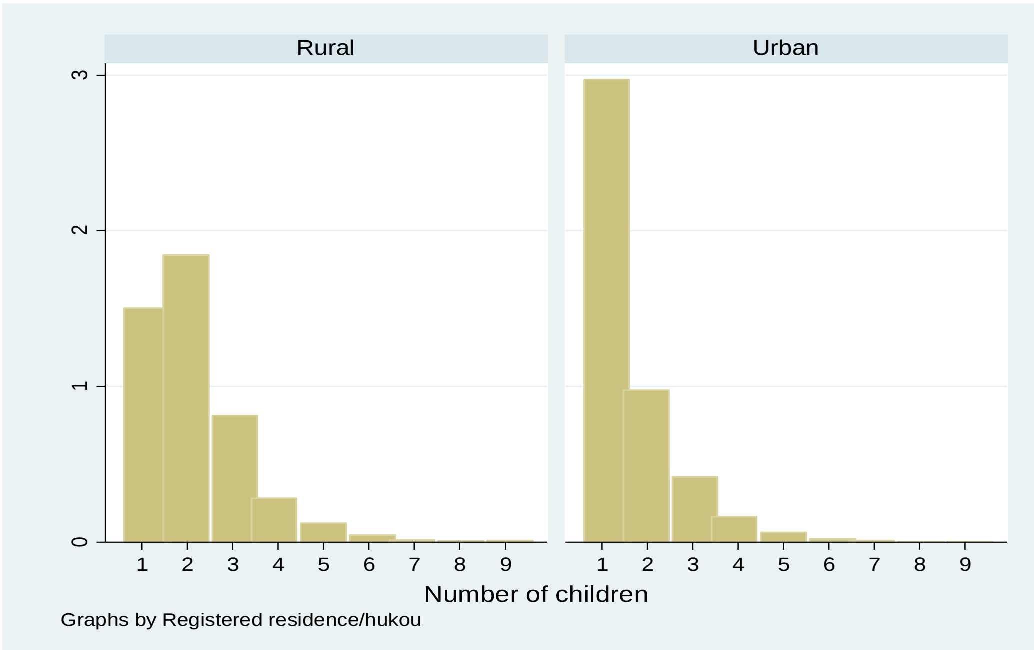


Figure 3.2 The Relation between the Sex Ratio and the Number of Children

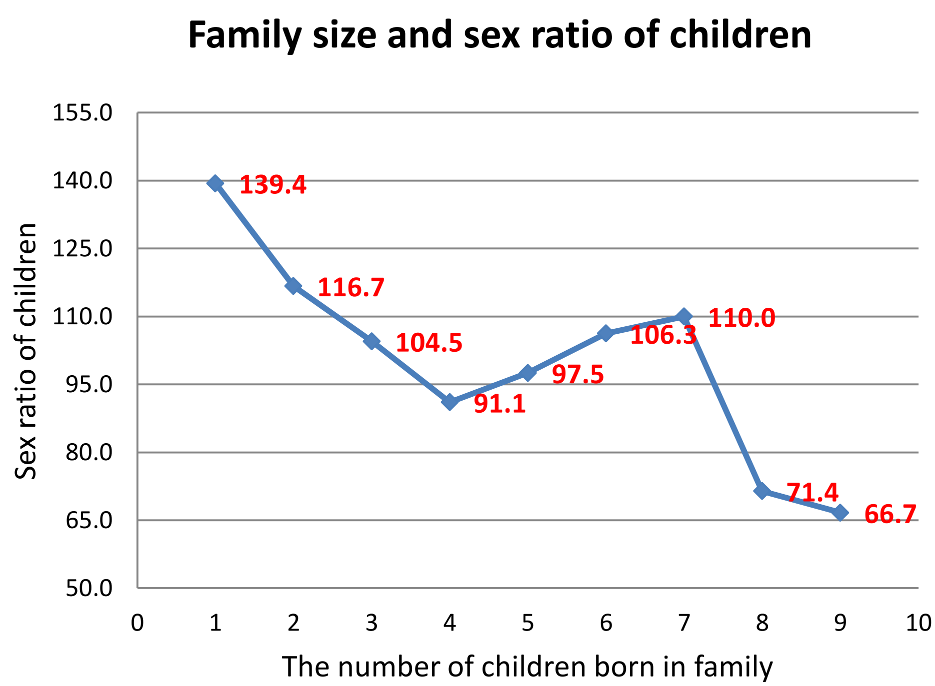
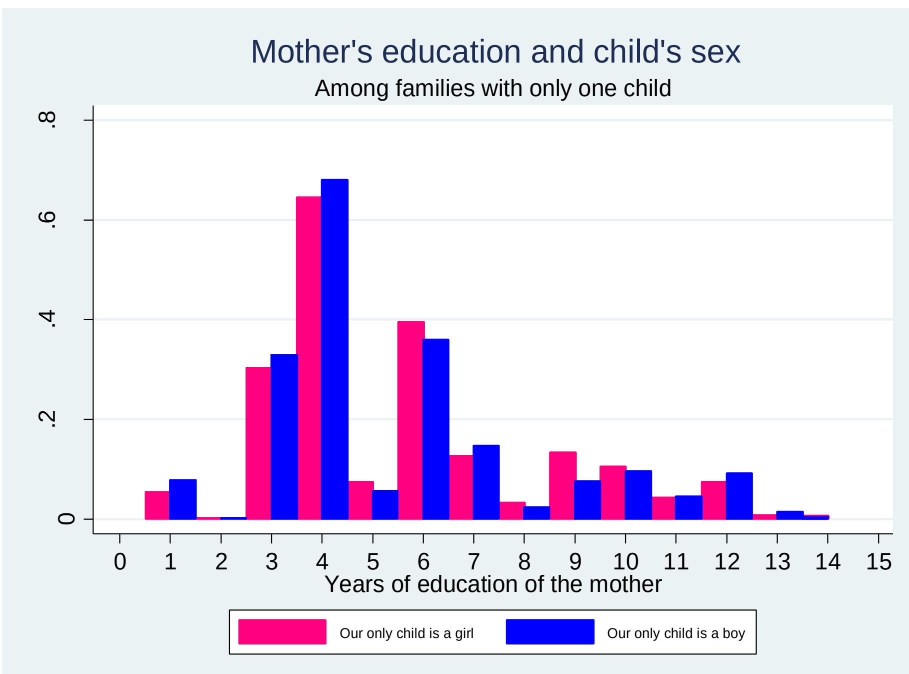


Figure 3.3 shows the relationship between mother's education and the sex composition of the children, illustrated with observations on the families with only one child. Among mothers who have not finished the 6-year primary school education, more of them tend to

have a boy as their only child. The 5-year education is a "watershed" because among mothers with 5 or more years of education, more of them tend to have a girl as their only child.

On the right tail we observe an incipient tendency that higher education is associated with higher frequencies of having a boy. For mothers with 11 or 13 years of education, the difference is very small. But according to our sample data, one point deserves to be mentioned here: mothers with 12-year education have clearly higher frequency of having a boy as the only child. There can be many interpretations for this. One of them is that higher education means better knowledge and accessibility to the sex manipulation resources, and therefore more of them end up with a boy as they desired for their only child. It is true if they truly prefer son over girl. But there is no evidence for that women with higher education do have such preference. And because observations with higher education constitute only 4% of only-children's parents, the frequencies of them having a boy or a girl may not be representative of the total level.

Figure 3.3 Mother's Education and the Sex Composition of Children



3.3 Sex composition of the Children

The most interesting variable is the "Boys among Children" (BAC) in a family. The sex ratio at birth (SRB) is computed as the ratio of male births to female births within a given region and a time interval. Theoretically we can have unlimited number of possible values for SRB. And we indeed observe large SRB disparity across regions and different socioeconomic groups. And this is what we focused on in the previous chapters. However, in the final analysis, it is not the provinces or counties that "create" the next generation; it is the parents or the whole family if it is an extended family. In this sense, we can consider the BAC as the SRB that is embodied in individual families. So it is of great importance to study how the BAC is determined within a family if we want to study birth masculinity of China.

The overall sex ratio of children in our CGSS sample is 114.76, with 4921 boys to 4288 girls. This is much lower than SRB for recent years. The main reason is BAC covers children of many ages, hence it is weighted average of historical SRB values (disregarded child mortality). It actually is close to China's SRB around 1995 which was 115(Bhattacharjya, Sudarshan, Tuljapurkar, Shachter, & Feldman, 2008), 13 years before the survey. Table 3.1 below shows how sex compositions of children across individual families contribute to the overall ratio between boys and girls. The observations are first divided into 9 groups by family size/the number of children in the household. Families with the same number of children are in the same group. Then we count the frequency for each form of sex composition conditioned on the family size. After that we can sum up the number of boys and girls in that group. Finally we work out the "sex ratio of children for each family size", which is simply the sum of boys divided by the sum of girls.

Take family size = 2 as an example. In our sample there are 1491 families who have two children. 886 of them have a boy and a girl, 245 have two girls and 360 have two boys. So the total number of girls born in these families is 245 times 2 plus 886, which is 1376. The total number of boys is 360 times plus 886, which is 1606. And the overall sex ratio of children among families with two children, is 1606 divided by 1376, which is 116.72.

Table 3.1 The Link of Sex Composition of Children and the Overall Sex Ratios

Boys Among Children	Freq.	Sum of Boys	Sum of Girls	Sex ratio of Children for each family size
0	1,048			(1)
1	1,461	1461	1048	139.41
0	245			
.5	886			(2)
1	360	1606	1376	116.72
0	43			
.3333333	291			
.6666667	241			(3)
1	74	995	952	104.52
0	10			
.25	74			
.5	90			
.75	50			(4)
1	11	448	492	91.06
0	4			
.2	19			
.4	25			
.6	25			
.8	22			(5)
1	1	237	243	97.53
0	1			
.1666667	3			
.3333333	6			
.5	11			
.6666667	7			
.8333333	4			(6)
1	1	102	96	106.25
.2857143	2			
.4285714	2			
.5714286	6			(7)
.7142857	2	44	40	110.00
.25	1			
.375	1			(8)
.625	1	10	14	71.43
.1111111	1			
.2222222	1			
.3333333	1			(9)
.6666667	2	18	27	66.67

3.4 The Model

Dependent variable

We want to examine how characteristics of individuals and their households affect the sex composition of their children. “Boys among Children” (BAC) can indicate the sex composition of the children, and thus is the variable we want to explain.

As we can see in Table 3.1, “Boys among Children” (BAC) is a discrete variable. Unlike the “sex ratio at birth” (SRB) we discussed in previous chapters, BAC cannot take unlimited number of values due to the biological or socioeconomic restriction on the number of children a couple can have in their entire life.

If there is only one child, it is either a boy or a girl. If a couple has two children, there are only three possibilities for their children sex composition: both are girls, both are boys or a boy and a girl. For a family having as many as 9 children, the forms of sex composition are more various but still limited. It is like the couple is facing with a group of options when they decide to have a certain number of children, and the actual BAC is the result of their decision and action. The question we are interested in here, is which factors lead the couple to end up with “this” form of sex ratio of children, instead of the other forms.

The largest family size in our sample is 9. With family size from 1 to 9, there are 29 options of sex composition of children:

(0, .11, .13, .14, .17, .20, .22, .25, .29, .33, .38, .40, .43, .44, .50, .56, .57, .60, .63, .67, .71 , .75, .78, .80, .83, .86, .88, .89, 1.00)

So there is a clear ordering of the outcome variable BAC. The size of difference is inconsistent, i.e. the spacing between successive categories is not even. For example, the difference between the first option 0 and the second option .11 is 0.11; while the difference between the second and the third option is 0.02. Therefore, BAC is not a continuous variable, and ordinary least squares (OLS) estimators are biased. The Ordered Logistic Regression will be a good choice.

More specifically, the General Ordered Response model will be used. One of the assumptions underlying ordered logistic regression is that the relationship between each pair of outcome groups is the same. In other words, ordered logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. And our primary analysis shows that our model violated this assumption. Therefore, we have to use the general ordered response model (Boes, 2006).

But we do not use BAC directly as our dependent variable. Instead we convert BAC into a real categorical variable. I generate a new variable called “dsp” with the label as “degree of son preference”, and categorize the original BAC values into five groups. dsp=1 means least son preference, dsp=5 means most son preference, while dsp=3 when BAC=0.5 which means balanced sex composition of children. Note that “degree of son preference” is not an attitudinal variable here. It is just the categorized boys’ number among children. It can be considered as the manifestation of the underlying son preference. And it is not a specific “son preference” from father or mother, but a combined or integrated “son preference” from both father and mother.

Explanatory variables

Based on the findings from the previous chapter, we assume that wife’s education (denoted as “wedu”) has a negative effect on the “degree of son preference”, while the husband’s education (“hedu”) has the opposite effect. And we assume that current employment (“ce”) of the wife is also relevant because if the wife has a decent job and relatively independent from the husband, then the family will show more respect to female. In such family, preference for son may not be very strong. We assume that lower fertility (number of children, “noc”) is associated with relatively more male births. Therefore, families with fewer children, the children are more likely are boys than girls. And we assume, ethnic groups (“minority”) do not have son preference or their son preference is not as strong as Han people. Urban couples (“urban”) usually have better access to sex-selective technology, but they are also more exposed to new values like gender equality thus have less preference for son than the rural couples. Therefore, we assume that urban families have lower son preference. Similarly, we consider the total family income ("tfinc") and the ownership of the house ("ownh") have negative effect on son preference because it means the parents have good resources to support themselves when they are older and do not rely on family support. This lowers the traditional value of sons as a parents' supporter. So parents who are wealthier have lower son preference. The last variable is the happiness of the wife ("hpns"). It is a self-evaluation variable from a question asked in the Questionnaire A of CGSS 2008. We assume that a happy wife has a lower preference for sons. She is happy about her life means she feels good about herself and thus is more likely to hold an equal gender norm.

Table 3.2 Variable Descriptions with CGSS Data

Variables	Descriptions
dsp	Degress of son preference(1= least preference, 5=most preference)
hedu	Education level of the husband (1=lowest, 6=highest)
wedu	Education level of the wife (1=lowest, 6=highest)
ce	Current employment of the wife (0=unemployed, 1=employed)
noc	Number of children in the family
minority	Ethnic group of the wife(0=Han, 1=Non-Han)
urban	Registered residence/hukou of the wife (0= rural, 1= urban)
tfinc	Total family income
ownh	Ownership of house (0=no, 1=yes)
hpns	Degree of happiness of the wife (1=least happy, 2=ok, 3=very happy)

Table 3.3 Summary of Variables with CGSS Data

Variable	Obs	Mean	Std. Dev.	Min	Max
dsp	5033	3.208027	1.644304	1	5
hedu	5497	3.312352	1.240883	1	6
wedu	5487	2.97649	1.313297	1	6
ce	6000	.6578333	.4744746	0	1
noc	6000	1.534833	1.20514	0	9
minority	6000	.0715	.2576797	0	1
urban	6000	.569	.4952574	0	1
tfinc	6000	607209.7	2327948	0	9999999
ownh	6000	.8105	.3919379	0	1
hpns	6000	2.0945	.5632844	1	3

Model specification:

The dependent variable is "degree of son preference", denoted as "dsp". It has 5 categories.

$$dsp_i = \begin{cases} 1, & Strong\ preference\ for\ daughters \\ 2, & Some\ preference\ for\ daughters \\ 3, & Balanced\ sex\ composition\ is\ preferred \\ 4, & Some\ preference\ for\ sons \\ 5, & Strong\ preference\ for\ sons \end{cases}$$

$$\Pr(dsp_i \le j|\boldsymbol{X_i}) = \boldsymbol{\Phi}(\boldsymbol{\alpha_j} - \boldsymbol{X_i'\beta}), j = 1,2,3,4,5$$

Where $\boldsymbol{X_i}$ is a 10×1 vector including the 9 explanatory variables (see Table 3.2) and an element 1to carry the intercept. $\boldsymbol{\beta}$ is also a 10×1 vector with elements as the 9 coefficients for the explanatory variables. And the $\boldsymbol{\alpha_j}$ are the threshold parameters, where $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_5 = \infty$.

We first run General Ordered Logit regression with the whole sample. Of course, observations without children are automatically excluded of course, left with 4720 observations. Then we analyze the model again with a subsample which contains the observations under the age of 49. We pay special interests in this group because they are still in the reproductive range or have just finished when the survey was conducted. The information on these children is the most recent information for us. This subgroup contains 2906 observations.

3.5 The Results

Table 3.4 Generalized Ordered Logit Regression with CGSS data (whole sample)

```
Generalized Ordered Logit Estimates                 Number of obs   =       4720
                                                    LR chi2(36)     =    9317.93
                                                    Prob > chi2     =     0.0000
Log likelihood = -2091.1093                         Pseudo R2       =     0.6902

------------------------------------------------------------------------------
         dsp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
        hedu |    .057857   .1275196     0.45   0.650    -.1920768    .3077907
```

wedu	-.1750185	.1264165	-1.38	0.166	-.4227903	.0727532
ce	.2533283	.2482309	1.02	0.307	-.2331952	.7398519
noc	5.443885	.2728422	19.95	0.000	4.909124	5.978645
minority	-.1569149	.3840356	-0.41	0.683	-.9096107	.595781
urban	-.0171322	.2580085	-0.07	0.947	-.5228196	.4885552
tfinc	-1.27e-08	5.55e-08	-0.23	0.819	-1.21e-07	9.61e-08
ownh	.1561165	.3578478	0.44	0.663	-.5452522	.8574853
hpns	-.4022137	.1918707	-2.10	0.036	-.7782734	-.0261541
_cons	-12.86494	1.342379	-9.58	0.000	-15.49595	-10.23392
2						
hedu	.0191627	.0595426	0.32	0.748	-.0975387	.1358641
wedu	-.1402631	.0677558	-2.07	0.038	-.2730621	-.0074642
ce	-.1017204	.1192397	-0.85	0.394	-.3354259	.1319851
noc	-.4579874	.0512968	-8.93	0.000	-.5585272	-.3574476
minority	.1437919	.1788315	0.80	0.421	-.2067114	.4942952
urban	.2190482	.1217299	1.80	0.072	-.019538	.4576344
tfinc	-9.45e-09	2.74e-08	-0.34	0.731	-6.32e-08	4.43e-08
ownh	.1251234	.1809641	0.69	0.489	-.2295596	.4798064
hpns	-.0442992	.0860917	-0.51	0.607	-.2130358	.1244374
_cons	2.281998	.5747816	3.97	0.000	1.155447	3.408549
3						
hedu	.0180571	.0607562	0.30	0.766	-.1010228	.137137
wedu	-.2788094	.0728935	-3.82	0.000	-.421678	-.1359408
ce	-.0399112	.1209408	-0.33	0.741	-.2769508	.1971284
noc	.2734288	.0507609	5.39	0.000	.1739393	.3729183
minority	.0221726	.1789203	0.12	0.901	-.3285047	.3728498
urban	.2578887	.1235283	2.09	0.037	.0157777	.4999997
tfinc	5.37e-08	2.59e-08	2.07	0.038	2.93e-09	1.05e-07
ownh	.1615075	.2021474	0.80	0.424	-.234694	.5577091
hpns	-.0841765	.0876522	-0.96	0.337	-.2559716	.0876186
_cons	-2.024779	.6035389	-3.35	0.001	-3.207693	-.8418644
4						
hedu	.1984378	.101822	1.95	0.051	-.0011296	.3980052
wedu	-.1898765	.1099795	-1.73	0.084	-.4054324	.0256795
ce	-.022393	.2351249	-0.10	0.924	-.4832293	.4384433
noc	-5.962442	.2965991	-20.10	0.000	-6.543766	-5.381118
minority	.3290429	.3181793	1.03	0.301	-.294577	.9526629
urban	.1275851	.258314	0.49	0.621	-.3787011	.6338712
tfinc	9.76e-08	3.78e-08	2.58	0.010	2.35e-08	1.72e-07
ownh	.2453205	.3467936	0.71	0.479	-.4343824	.9250234
hpns	-.4325719	.1806954	-2.39	0.017	-.7867284	-.0784155
_cons	14.92473	1.332781	11.20	0.000	12.31253	17.53693

We find that higher level of wife's education is associated with a lower son preference, in both the second and third categories of degree of son preference. The formal interpretation is for one unit increase in wife's education, we expect a 0.14 decrease and a 0.27 decrease in the log odds of preferring son over daughter, given all of the other variables in the model are held constant, in the second and the third category of "dsp", respectively. But husband's education is not statistically significant for the determination of the degree of son preference.Besides, the number of children is also significant. We find that it has positive effect on the odd orders of “dsp”. That is the number of children does not have a stable effect on the son preference. And the higher level of happiness of the wife seems to be associated with a lower son preference. The total family income is also a weak explanatory variable. It means the general well-being of the family is not relevant cannot indicate a degree of son preference in that family.

Table 3.5 Generalized Ordered Logit Regression with CGSS data (sub-sample)

```
Generalized Ordered Logit Estimates                Number of obs   =       2906
                                                   LR chi2(36)     =    6287.81
                                                   Prob > chi2     =     0.0000
Log likelihood = -498.06571                        Pseudo R2       =     0.8632
```

dsp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
hedu	.2985238	.3015704	0.99	0.322	-.2925433	.8895909
wedu	-.3996359	.2495003	-1.60	0.109	-.8886474	.0893757
ce	.6586929	.5026689	1.31	0.190	-.3265201	1.643906
noc	6.520621	.5931522	10.99	0.000	5.358064	7.683178
minority	-.2198773	.801844	-0.27	0.784	-1.791463	1.351708
urban	-.5515582	.5196258	-1.06	0.288	-1.570006	.4668895
tfinc	-1.33e-07	1.12e-07	-1.19	0.235	-3.52e-07	8.64e-08
ownh	-.6273242	.6415572	-0.98	0.328	-1.884753	.6301049
hpns	-.8340052	.3710404	-2.25	0.025	-1.561231	-.1067795
_cons	-14.07764	2.550424	-5.52	0.000	-19.07638	-9.078897
2						
hedu	-.0977455	.1411588	-0.69	0.489	-.3744116	.1789206
wedu	-.2588701	.1431572	-1.81	0.071	-.5394529	.0217128
ce	-.3449672	.3140396	-1.10	0.272	-.9604735	.2705391
noc	-2.118402	.1784994	-11.87	0.000	-2.468255	-1.76855
minority	-.0295828	.3458228	-0.09	0.932	-.707383	.6482174
urban	.3527962	.2935683	1.20	0.229	-.222587	.9281795
tfinc	-1.61e-08	8.37e-08	-0.19	0.847	-1.80e-07	1.48e-07
ownh	-.2644894	.3188834	-0.83	0.407	-.8894895	.3605106
hpns	-.4008528	.1939829	-2.07	0.039	-.7810524	-.0206532
_cons	9.33436	1.275386	7.32	0.000	6.83465	11.83407
3						
hedu	.3598422	.1703479	2.11	0.035	.0259663	.693718
wedu	-.7925853	.1816763	-4.36	0.000	-1.148664	-.4365063
ce	-.1160696	.345846	-0.34	0.737	-.7939153	.561776
noc	.8266231	.176913	4.67	0.000	.47988	1.173366
minority	.2397129	.3819207	0.63	0.530	-.508838	.9882638
urban	.8816439	.3221615	2.74	0.006	.250219	1.513069
tfinc	1.87e-07	6.54e-08	2.87	0.004	5.92e-08	3.15e-07
ownh	.2913873	.354563	0.82	0.411	-.4035434	.986318
hpns	.5270684	.2424601	2.17	0.030	.0518554	1.002281
_cons	-6.665837	1.488033	-4.48	0.000	-9.582328	-3.749345
4						
hedu	.4617592	.3581249	1.29	0.197	-.2401528	1.163671
wedu	-.7708926	.3518521	-2.19	0.028	-1.46051	-.0812751
ce	.1647593	.9091367	0.18	0.856	-1.617116	1.946634
noc	-9.799892	1.246622	-7.86	0.000	-12.24323	-7.356558
minority	.1623617	.773323	0.21	0.834	-1.353324	1.678047

```
       urban |  -1.016763   1.184203    -0.86   0.391    -3.337758    1.304232
       tfinc |  -1.54e-07   3.09e-07    -0.50   0.617    -7.59e-07    4.51e-07
        ownh |   .0087282   .7165922     0.01   0.990    -1.395767    1.413223
        hpns |   .1160024   .5166984     0.22   0.822     -.896708    1.128713
       _cons |   26.29657   5.376498     4.89   0.000     15.75883    36.83431
------------------------------------------------------------------------------
```

We also observe a quite complex effect of the fertility level on the degree of son preference. But we observe a significant effect of wife's education on the decline of son preference. We find for one unit increase in wife's education, we expect a 0.79 decrease and a 0.77 decrease in the log odds of preferring son over daughter, given all of the other variables in the model are held constant, in the third and fourth categories of "dsp". Wife's education shows a much stronger effect compared with that suggested by the whole sample. This means, within the younger group, wife's education is significant and strong factor for the decline in son preference. On the other hand, we observe a positive relation between husband's education and the degree of son preference. Our data suggests that with one unit of increase in husband's education, the log odds of preferring son over daughter is expected to increase by about 0.36, while the other variables are held constant.

3.6 Discussion

Analyses with the individual level data confirm some of our finding from the macro level models. We find higher level of wife's education is associated with lower level of son preference. It seem like higher education among women makes individuals more open to modern gender norm and abandon son preference, which results in fewer sex-selections and thus reduce the overall SRB.

But the fertility or the number of children is not a stable factor for the degree of son preference. The effects are significant but with opposite signs in different cases. It corresponds to the result from aggregate data analyses: low total fertility rate is not the main reason for inordinate SRB.

And family's wealth is also a weak explanatory variable for the degree of son preference. And the general well-being of the parents is not relevant and cannot indicate a degree of son preference in that family. These correspond to the finding with aggregate data: socioeconomic status does not explain much SRB disparities across regions.

To sum up, with the individual data, we can still identify a strong effect of education on the degree of son preference, while other explanatory variables are not significant suggested by our data.

4 Conclusions

In this thesis, we collected data on both provincial and county level from the 2010 census data in China. We studied the geographic disparities in China of the Sex Ratio at Birth (SRB), defined as the number of baby boys per 100 baby girls born during a certain period. We also analyzed the determinants for the regional differentials of SRB, with a focus on the education's effects. Then we used survey data from CGSS 2008 to analyze how sex composition is determined by individuals and households characteristics. The main conclusions can be summarized as following.

(1) Education has significant effects on SRBs, suggested by both the macro level data and the individual data. Improvement on people's education can substantially reduce SRB, mainly through the reduction on son preference. Moreover, the female's relative education has significantly large effect on the decline in SRBs. In view of the main education variable we used in the aggregate analyses, "ays" comprises the whole population, including people beyond reproductive age span. Ideally I would have like to include "ays" for persons of prime reproductive ages only, but that data was not available from the 2010 census tabulations.

(2) The socioeconomic status is not a very important factor for SRB. Using the latest census data we could not identify a clear effect of the socioeconomic status on SRB. The inverted "U" relation suggested by Guilmoto and Ren (2011) has not been identified by our studies.

(3) We find from the provincial data that SRB is increasing in fertility. But results from the county-level data show that it is no longer significant while other variables are introduced into the model. And with the CGSS data, we find the number of children is not a stable factor for the degree of son preference. Therefore, we conclude lower fertility is not the main reason for high SRB.

(4) Due to lack of public resources of support for the elderly, the traditional elderly support form is still prevailed. It reinforces people's preference for sons and tends to increase SRB.

Contrary to the earlier findings which suggest higher education gives better access to sex-selective abortion and thus increases the excess male births, this thesis finds a clear effect of education on decreasing the level of SRB. At the individual level, increase in mothers' education is going to reduce the number of boys among the children. Higher education leads to a better well-being of women, winning respect for females in general. So the family will be more open to modern gender norm and abandon son preference, which results in fewer sex-selections and thus reduce the overall SRB. The effect of education on aggregate level suggests if there are quite a few people who are highly educated and start to adopt new gender norm. And such norm can be spread fast through a society, then the effects of education on SRB are actually amplified by passing the new gender norm to the less educated group.

The main inspiration from this thesis is in current China, education has started to play a role of reducing people's preference for sons. The improvement in people's education level, especially the female education level, will accelerate the pace for sex ratio at birth to return to the normal range.

Reference

Attané, I. (2009). The determinants of discrimination against daughters in China: Evidence from a provincial-level analysis. *Population Studies, 63*(1), 87-102.

Banister, J. (2004). Shortage of girls in China today. *Journal of Population Research, 21*(1), 19-45.

Bhattacharjya, D., Sudarshan, A., Tuljapurkar, S., Shachter, R., & Feldman, M. (2008). How can economic schemes curtail the increasing sex ratio at birth in China? *Demographic Research, 19*(54), 1831.

CGSS. (2008). Chinese General Social Survey.

Chen, Y.-h., & Hu, X.-w. (2012). Social Change and the Turning Point of Sex Ratio at Birth. *Population & Development, 18*(1).

Chu, L.-W., & Chi, I. (2008). Nursing homes in China. *Journal of the American Medical Directors Association, 9*(4), 237-243.

Chung, W., & Das Gupta, M. (2007). The decline of son preference in South Korea: The roles of development and public policy. *Population and Development Review, 33*(4), 757-783.

Croll, E. (2002). *Endangered daughters: Discrimination and development in Asia*: Routledge.

Drèze, J., & Murthi, M. (2001). Fertility, education, and development: evidence from India. *Population and Development Review, 27*(1), 33-63.

Ebenstein, A. (2010). The "missing girls" of China and the unintended consequences of the one child policy. *Journal of Human Resources, 45*(1), 87-115.

Filmer, D., Friedman, J., & Schady, N. (2009). Development, modernization, and childbearing: The role of family sex composition. *The World Bank Economic Review, 23*(3), 371-398.

Fuse, K. (2010). Variations in attitudinal gender preferences for children across 50 less-developed countries. *Demographic Research, 23*(36), 1031-1048.

Goodkind, D. (2011). Child underreporting, fertility, and sex ratio imbalance in China. *Demography, 48*(1), 291-316.

Gu, Wang, Guo, & Zhang. (2007). China's local and national fertility policies at the end of the twentieth century. *Population and Development Review, 33*(1), 129-148.

Gu, B., & Li, Y. (1994). *Sex ratio at birth and son preference in China.* Paper presented at the UNFPA Symposium on Issues Related to Sex Preference for Children in the Rapidly Changing Demographic Dynamics in Asia, Seoul, November.

Gu, B., & Roy, K. (1995). Sex ratio at birth in China, with reference to other areas in East Asia: what we know. *Asia Pacific Population Journal, 10*, 17-42.

Guilmoto, & Ren. (2011). Socio-economic Differentials in Birth Masculinity in China. *Development and Change, 42*(5), 28.

Guilmoto, C. Z. (2008). Economic, social and spatial dimensions of India's excess child masculinity. *Population (english edition), 63*(1), 91-117.

Guilmoto, C. Z. (2009). The Sex Ratio Transition in Asia. *Population and Development Review, 35*(3), 519-539.

Guo, Z. G. (2007). Multilevel Analysis of Sex Ratio at Birth in China 2000 Population Census. *RenKouYanjiu, 3*, 002.

Hesketh, T. (2011). Selecting sex: The effect of preferring sons. *Early human development, 87*(11), 759-761.

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2008). *Principles of econometrics* (Vol. 5): Wiley Hoboken, NJ.

Jha, P., Kesler, M. A., Kumar, R., Ram, F., Ram, U., Aleksandrowicz, L., . . . Banthia, J. K. (2011). Trends in selective abortions of girls in India: analysis of nationally representative birth histories from 1990 to 2005 and census data from 1991 to 2011. *The Lancet, 377*(9781), 1921-1928.

Jolliffe, I. (2005). *Principal component analysis*. http://f3.tiera.ru/2/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf: Wiley Online Library.

Li, H., Yi, J., & Zhang, J. (2011). Estimating the effect of the one-child policy on the sex ratio imbalance in China: identification based on the difference-in-differences. *Demography, 48*(4), 1535-1557.

NSB. (2010). (National Statistics Bureau)Tabulation on the 2010 Population Census of the People's Republic of China.

Retherford, R. D., & Roy, T. (2003). Factors affecting sex-selective abortion in India and 17 major states.

Steel, D. G., Tranmer, M., & Holt, D. (2006). Unravelling ecological analysis. *Advances in Decision Sciences, 2006*.

Tan, L. Y. (2008). Changing mindsets: How China's abnormal sex ratio is turning its government into a champion of gender equality. *Asian Journal of Comparative Law, 3*(1).

Yang, J. (2006). Reginal Diversity of Fertility and Child Sex Ratio in China. *RenKouYanjiu (Chinese), 30*(3).

Yanjie, B., & Lulu, L. (2008). *China General Social Survey (China GSS)*. Retrieved from: http://doi.org/10.3886/ICPSR23861.v1

Zeng, Y., Tu, P., Gu, B., Xu, Y., Li, B., & Li, Y. (1993). Causes and implications of the recent increase in the reported sex ratio at birth in China. *Population and Development Review*, 283-302.

Zhai, Z.-w., & Yang, F. (2009). The level of sex ratio at birth of China and the data quality *Population Journal (Chinese), 176*(4).

Zhu, W. X., Lu, L., & Hesketh, T. (2009). China's excess males, sex selective abortion, and one child policy: analysis of data from 2005 national intercensus survey. *BMJ: British Medical Journal (Overseas & Retired Doctors Edition), 338*(7700).