STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
659 Degree Project in Economics
Spring 2022

Gender Imbalance and Educational Investments in China

Jessica Johansson (24817) and Anna Johnsson (24630)

Abstract: An unintended consequence of the One Child Policy, introduced in 1979, has been the increasing sex ratio imbalance in China. This has resulted in millions of excess males, facing difficulties in finding a spouse. The increased marriage market competition has caused behavioral changes among males, motivated by an eagerness to increase their attractiveness towards potential partners. This paper seeks to examine whether individuals facing fiercer marriage market competition attain additional education as a strategy to improve their attractiveness. Human Capital Theory stipulates that education increases an individual's life-time earnings, thus we hypothesize that males faced with increased marriage market competition may use education as a strategy to acquire wealth and signal high-earnings potentials, as wealth has been proven to be an important determinant of an individual's marriage market attractiveness. Using data from 2010 China Family Panel Studies and 1990 and 2000 China Population Censuses, we investigate this relationship by using an Ordinary Least Squares (OLS) and Instrumental Variable (IV) approach. We found support for our hypothesis that higher gender imbalance results in increased educational attainment for males, although the economic significance is not excessively high. Surprisingly, we also found that the educational effect of the gender imbalance is in fact stronger for females than males. Similar research has been made in the 20th century US immigrant setting, but we contribute to existing research by extending the knowledge of the relationship to a Chinese 21st century context, a geographical and temporal research space which has so until now been unexplored.

Keywords: Sex Ratio, Education, One Child Policy, Marriage, Signaling

JEL: D31, I18, I24, I26, J12, J13, J16, J24

Supervisor: Céline Zipfel
Date submitted: May 16, 2022
Date examined: May 24, 2022
Discussants: Cajsa Dalne and Gustav Ring
Examiner: Johanna Wallenius

Acknowledgements

We would like to express our gratitude to our supervisor Céline Zipfel for guidance and support throughout the process. Further, we would also like to give a special thanks to Professor Xiaobo Zhang at the University of Peking for helping us with data access.

Disclaimer

This study has been written from a heteronormative perspective due to the nature of the research topic. Hence, we do not take other sexual orientations into account, although they certainly exist in our sample. Further, as neither same-sex marriages not civil unions are recognized by Chinese laws, we make the simplification of a heterosexual norm in our investigation (Foreign Language Press, 1982).

Definitions & Abbreviations

Asymmetric information: when one party in a transaction has more information than the other.
Confucianism: an ancient Chinese philosophy, with a special focus on ethics and morality.
Malthusian catastrophe: when a growing population surpasses the planet's food production capacity.
Patriarchy: a social system of male dominance, upheld by structures favoring men.
Patrilineage: when descent is traced through the male line.
Patrilocality: when a married couple lives with or close to the husband's family.
Sex Ratio at Birth: the number of boys born alive per 100 girls born alive during a specific period of time.

IV: Instrumental Variable
OCP: One Child Policy
OLS: Ordinary Least Squares

Table of contents

1. Introduction

In 1979, the People's Republic of China introduced several economic reforms. From that period and onwards, China has been one of the world's fastest growing economies. Between 1979 and 2018, China's Gross Domestic Product (GDP) grew at an impressive average compound annual rate of 9.50 percent with real per capita disposable income more than tripling over the period (Morrison, 2019; Oliveira, 2018). According to the World Bank, China is currently the second largest economy in the world, with a GDP of $14.72 trillion in 2020 (World Bank, 2022). The rapid development has lifted over 800 million people out of poverty and has resulted in significant improvements in access to health, education and also fueled a growing private labor market (World Bank, 2022; Morrison, 2019; Yee, 2006).

In terms of demographics, China is one of the world's most populous countries with more than 1.3 billion citizens in 2010 and over 1.4 billion in 2021(World Bank, 2022). The rapidly growing population was one of the main motives underlying the implementation of family planning policies in the 1970's. The One Child Policy (OCP), introduced 1979, allowed Chinese parents to only have one child (although moderations were made over time) and has been considered the world's strictest family planning policy ever implemented (Gietel-Basten et al., 2019). An unintended consequence of the family planning policies has been the rising sex ratio imbalance, often attributed to a combination of a cultural son preference, increased access to prenatal determination tools and the imposed fertility limitations (Zhang and Wei, 2011a). Prior to the OCP in 1979, China had a sex ratio at birth of 107 males per 100 females. In 1990, the sex ratio at birth had increased to 112 males per 100 females, reaching its peak of approximately 117 males per 100 females in 2005 (United Nations, 2019)[1]. This demographic development has been of interest to many scholars trying to establish the consequences of the rising gender imbalance (Ebenstein, 2010; Sen, 1990, Yi et al, 1993). Nobel prize laureate Amartya Sen estimated that the neglect of girls resulted in over 50 million women "missing" in China in the 1990's, and the situation has not improved materially since then (Sen, 1990). In 2020, China displayed a factual surplus of approximately 35 million males (China statistical yearbook, 2021). Out of these, 16 million were in what is considered a "marriageable age". The excess number of males has resulted in an increased marriage market competition, where millions of males risk not finding a spouse to marry.

A large branch of literature has pointed out several negative consequences of the imbalanced sex ratio, such as increased crime rates, mortality, and mental health issues among Chinese men (Zhang et al., 2021; Zhang et al., 2020; Edlund et al., 2013). Males and their families have also been found to engage in more entrepreneurial activities, increase their savings rate, and invest more in high-risk assets on the stock market in order to improve their financial status, as wealth is considered an important determinant of an individual's marriage market attractiveness (Chang et al., 2011; Becker, 1981). One topic that has not been thoroughly studied, and that this paper seeks to examine, is whether individuals facing fiercer marriage market competition also attain additional education. As Human Capital Theory stipulates that education increases an individual's life-time earnings, we hypothesize that males faced with increased marriage market competition, may use

[1] Estimates of sex ratio at birth vary in the research, where several studies have found the sex ratio at birth peaking at levels well above 120 males per 100 females, such as (Jiang & Zhang, 2021; Zhang, He & Ma, 2021).

education as a strategy to acquire additional wealth and as a method to signal high-earnings potentials on the marriage market. More specifically, we intend to answer:

How does a skewed sex ratio impact males' investments in education?

By utilizing data from the 2010 China Family Panel Study (CFPS) and the 1990 and 2000 China Population Censuses, we examined the relationship between sex ratio imbalance and completed years of education. The CFPS dataset is a national, cross-country survey collecting information about economic, non-economic and social aspects of the Chinese population at the individual, household, and community level in both rural and urban China (CFPS, 2022). The population censuses capture demographic, economic and social information about the Chinese population which enabled us to calculate the youth sex ratio in the individuals' province of birth (IPUMS). Through combining the CFPS and the Population Censuses we were able to empirically test whether a skewed sex ratio affects males' completed years education for the age group 16-35 in 2010.

To do so, we applied two empirical methods. Firstly, we utilized the simple Ordinary Least Squares (OLS) method, where years of education was regressed on the prevailing sex ratio during an individual's youth in their province of birth. Further, to control for potential omitted variable bias and other endogeneity issues, we applied an Instrumental Variable (IV) methodology, where financial penalties for violating the OCP were used as instruments for the sex ratio.

We found support for our hypothesis that higher gender imbalance results in increased educational attainment for males, although the economic significance is not excessively high. Surprisingly, we also found that the effect of the gender imbalance is in fact stronger for females than males. We are hypothesizing that these effects could stem from the prevailing son preference, disadvantaging females in the labor market, resulting in females attaining additional education in order to compete for the same labor opportunities as males. Further, the economic growth of the Chinese economy might also have given rise to a larger supply of high-income jobs, not requiring higher education. For males facing a competitive marriage market, these job opportunities might be an attractive alternative with short-term benefits, as the average marriage age is relatively low.

Our study contributes to the body of existing research in several ways. First, we extend the knowledge of the relationship between skewed sex ratios and educational attainment to a Chinese 21st century context, a geographical and temporal research space which currently is unexplored. Secondly, investigating this topic can be an important contribution from a gender equality point of view, as it can provide guidance to policy makers on factors affecting educational inequality between genders.

The remainder of the paper is organized as follows. Section two provides a background on the sex ratio development, marriage traditions and educational system in China. Section three presents the theoretical frameworks used in this study. Section four presents our literature review. Section five discusses the research gap, contribution to literature, research question and hypothesis. Section six provides the data sets used for our empirical investigation. Section seven provides our empirical

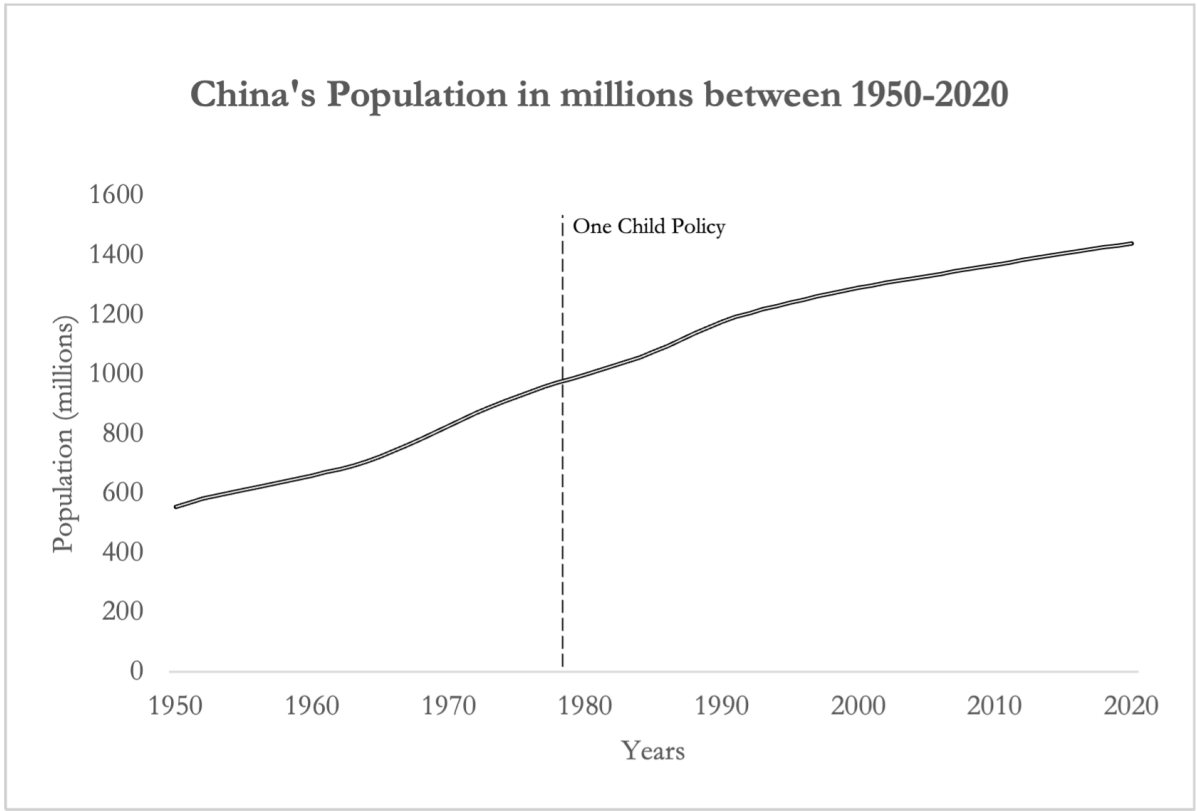
methods. Section eight presents our results. Section nine reports our discussion, limitations, and suggestions for future research. Finally, section ten provides our conclusions.

2. Background

2.1 Fertility limitations and the One Child Policy

In the 1970's, the Chinese government under the ruling of Mao Zedong, became increasingly concerned about the growing population (see figure 1). Mao's Communist Party feared a Malthusian catastrophe (see definitions & abbreviations), with insufficient resources for the increasingly large population, leading to extensive famine (Ebenstein, 2010). As a result, far-reaching family planning incentives were introduced in 1973 (Wang, 2012). These included extended contraceptive and abortion services into rural areas and governmental campaigns promoting later marriage, longer time intervals between children and smaller sized families under the slogan "Later, Longer, Fewer" (Wang, 2012). Despite these incentives, the population continued to grow throughout the 1970's and the measures were considered insufficient since China was not expected to meet its official population target of 1.2 billion people by the year 2000 (Gietel-Basten et al., 2019).

Figure 1: China's population growth between 1950-2020



Source: United Nations (2019)

Subsequently, in 1979, the central government under the leadership of Deng Xiaoping introduced the One Child Policy (OCP), a set of regulations that restricted Chinese families to having one child only (although ethnical minorities faced more relaxed restrictions) (Zhang, 2017; Wang, 2012). The government implemented several abortion and sterilization campaigns, and families violating the fertility limits were punished through financial fines, calculated as a percentage of the annual income (ranging between 0.1-5.0 times of the annual income). Furthermore, illicit children could be denied access to public health care and education, thus, they were not given the same opportunities as other children (Ebenstein, 2010; Oliveira, 2018).

Although the policy was implemented on a national level in 1980, the central government implemented further localized fertility policies in 1984. The majority of urban families were restricted to having one child, while many rural families were allowed two children if their first-born child was a girl (referred to as the "1.5 child policy") (Zhang, 2017). In some local communities, families living in remote areas were allowed 2-3 children (Oliveira, 2018; Ebenstein, 2010). The implementation of the localized policies resulted in a variety of fertility limits and financial penalties for violations among regions, provinces, and ethnic groups (Zhang, 2017).

The OCP was tightly enforced until 2015 when it was replaced by a Two-Child Policy. In May 2021, a Three-Child Policy was formally passed into law (BBC, 2021).

2.2 Son preference and the sex ratio development

One key characteristic of the Chinese culture is their prevailing preference for sons (Jiang and Zhang, 2021; Ebenstein, 2010; Das Gupta and Shuzhuo, 1999; Coale and Banister, 1996). Researchers attribute this phenomenon to China's Confucian heritage, cultural traditions, and large agricultural dependence (Guo et al., 2018; Jayachandran, 2015; Jiang et al., 2011). Sons are not only considered an important labor force in agricultural production, they are also relied on to support the older generation through patrilocality, remaining in the family after marriage (Jiang et al., 2011). Further, China's tradition of patrilineality, patrilocality and patriarchy systems (see definitions & abbreviations) have historically given males superior rights in wealth inheritance, living arrangements and family power structure, resulting in a higher status in the Chinese society (Jiang and Zhang, 2021; Jiang et al., 2011).

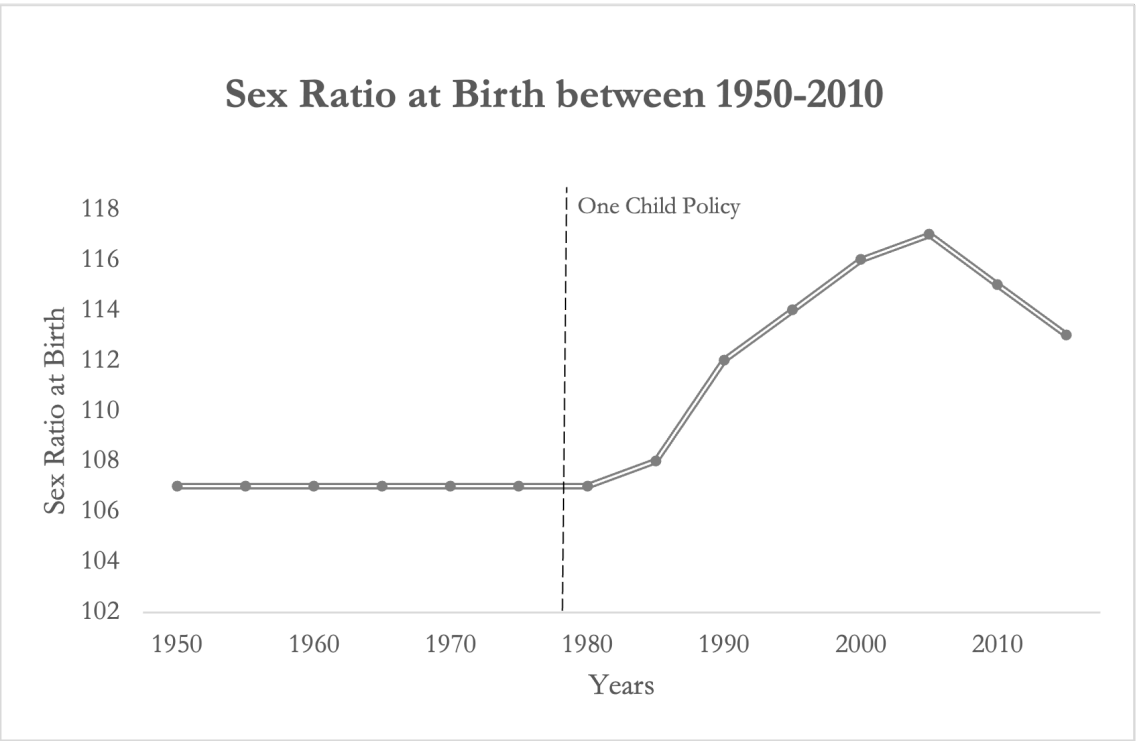
Even prior to the implementation of the different family planning policies, the preference for sons had resulted in skewed sex ratios at birth in China, although this was often materialized through families having additional children in the hope of receiving a son (Jiang and Zhang, 2021)[2]. The fertility limits imposed under the OCP resulted in couples engaging in sex selective abortions, especially after the introduction of inexpensive ultrasound in the 1980's (Oliveira, 2018; Ebenstein, 2010). Consequently, China has experienced a growing gender imbalance at birth, reaching its peak of approximately 117 boys per every 100 girls born in 2005 (see figure (2)) (Tang, 2021)[3]. Due to the different fertility limits imposed by local authorities, sex selective behaviors differed regionally,

[2] The expected sex ratio at birth in the absence of interference lies between 103 to 107 boys per 100 girls (Ritchie and Roser, 2019).

[3] Estimates of sex ratio at birth vary in the research, where several studies have found the sex ratio at birth peaking at levels well above 120 males per 100 females, such as (Jiang and Zhang, 2021; Zhang et al., 2021).

with provinces subject to the 1.5 child limit experiencing the highest gender imbalance growth (Jiang et al., 2011). This has resulted in significant variations in the sex ratio development across China.

Figure 2: Sex Ratio at Birth in China between 1950-2010, in five-year periods



Source: United Nations (2019)

In 2020, the Chinese population consisted of approximately 723.57 million males and 688.55 million females. This implies an excess of approximately 35 million males, of which approximately 16 million were in what is considered "marriageable age" (20-40 years old) (China Statistical Yearbook, 2021). One consequence of the imbalanced sex ratio is the difficulty for a substantial number of males to attract a female partner. Research has found that it is often the poor, uneducated males (commonly referred to as the "Guanggun" or "bare-branches") in rural areas that face these difficulties (Hesketh et al., 2011).

2.3 Marriage traditions in China

Family and marriage are central aspects of Chinese culture. As stated by Bai, Tang and Xie, "it is still a common notion today that every male should take a wife and every female should take a husband" (Bai et al., 2022). The minimum marriage age for women is 20 and for males 22, although later marriage is encouraged by the government (Foreign Language Press, 1982). However, it is not unusual that parents and relatives put pressure on unmarried sons and daughters to get married before the age of 30 (a phenomenon referred to as "Cui Hun" or "urge to marry"), especially for women (Tan and Xu, 2020).

It should be noted that China is not only the most populous country in the world but has also faced rapid development in recent decades, resulting in large differences in terms of culture, where certain areas display more westernized customs while other regions maintain more traditional values. However, most literature on this topic does typically cover traditional Chinese marriage ideas. It is thus important to note that the information presented in this section does not consider regional differences nor presents a collectively exhaustive description of Chinese marriage customs.

In traditional Chinese marriages, there are typically large costs associated with marriage for the male and his family. Firstly, it is custom that the groom's family finance the wedding ceremony, as well as pay a "bride price" (reversed dowry) to the bride's family. Sometimes, the bride's family reciprocates the tradition and pays a dowry back to the groom's family, although this custom is less common (Fu et al., 2018; Andersson, 2007). Further, many brides also expect their future husband to provide accommodation, which results in additional large expenditures for the groom's family, especially in recent years after spikes in house prices in China (Wei et al., 2017). Additionally, the continued sex ratio imbalance, has caused rising "bride prices", putting further economic pressure on the groom's family (Li and Li, 2021).

Traditionally, parents of the groom are expected to help their son find a bride, especially in rural areas where the patrilocality system dominates. This system makes sons and their wives the most important source of support for elderly parents, as rural areas often lack reliable old-age support systems (Chen and Silverstein, 2000). As stated by Jin, Guo, and Feldman "if parents are unable to arrange a marriage by providing bride price, betrothal gifts, and marriage ceremony expenses, and so on for a son, it could justify the son not supporting them [at old age]" (Jin et al., 2015). Further, as the marriage tradition outlined above is the norm in traditional Chinese culture, parents who fail the task will suffer disapproval and social pressure (Bai et al., 2022).

Although males and females in theory can marry freely, certain structures dictate which marriage candidates are considered appropriate[4]. One common division of the Chinese population is made by the prevailing household registration system, referred to as 'Hukou'. Until 2014, all Chinese citizens were born with either an agricultural or non-agricultural hukou status and a hukou location registration (Wang, 2021)[5]. Depending on Hukou status, citizens have different access to the welfare system, where a non-agricultural hukou has been considered advantageous in terms of social security, education, and employment opportunities (Wu and Treiman, 2004; Huang, 2001). Previous research has found that the hukou system has created a hierarchy of preferred hukou statuses, where a non-agricultural hukou status is associated with higher social standing (Wu and Treiman, 2004). Due to this fact, citizens typically marry within their own hukou-status, although recent relaxations of the system and the general development in China have opened up for more inter-hukou marriages (Huang, 2001).

[4] As opposed to same sex marriages which is not recognized in China (Foreign Language Press, 1982).
[5] The division between agricultural and non-agricultural citizens was abolished in 2014 and replaced by a national residency registration (jumin hukou) (Wang, 2021).

2.4 Educational system in China

The Chinese educational system contains primary, secondary, and higher education (see table (1)) (Britannica, 2022). Secondary education includes both junior and senior education, while higher education consists of senior vocational, bachelor's, master's, and PhD levels. Primary education comprises six years of education, while secondary education consists of three years of junior and senior school each. Higher educational programs vary in length but typically take between 2-4 years to complete.

The first nine years of schooling (primary and junior secondary school) are mandatory and paid for by the government under the "Compulsory Education Law" from 1986 (Britannica, 2022; Ministry of Education, 2009). After completing mandatory schooling, students can choose to proceed to senior secondary school, in either an academic or vocational secondary program (OECD, 2016).

Table 1: Educational levels in China

Age	Years of schooling	
27	22	PhD programme
26	21	
25	20	
24	19	Master's programme
23	18	
22	17	
21	16	University (bachelor's degree) and vocational college
20	15	
19	14	
18	13	
17	12	Senior secondary school
16	11	
15	10	
14	9	Junior secondary school
13	8	
12	7	
11	6	Primary school
10	5	
9	4	
8	3	
7	2	
6	1	
5		Pre-school and kindergarten
4		
3		

Source: OECD (2016)

The Chinese hukou system dictates access to public schooling, which disincentivizes parents of young children to move outside the province since the children will not unconditionally be given access to public schooling in their new area of residence (Zhang et al., 2021; Wang et al., 2017).

The Chinese educational system has undergone substantial transformations since the 1970's. During the ruling of Mao, the focus of the educational system shifted towards efficiency and growth (Gruijters et al., 2019). It was also during this time that the nine years of mandatory schooling were implemented. Since the 1980's, enrollment in education has increased substantially, with almost all individuals born after 1980's having attained the nine-year compulsory education (Zhang, 2022). Due to a shift in central policy during this period, school resources were decentralized, resulting in considerable differences in availability and quality of education among regions, especially between rural and urban areas (Gruijters et al., 2019). In 1990, there was a second wave of reforms, which focused on increasing access to higher levels of education. Further, the government implemented policies targeting equal resource distribution, access, and quality among schools (Huang et al., 2015). As a result, higher educational attainment rates have surged in recent years. In 2019, 28 percent of 18-year-olds, 39 percent of 19-year-olds and 40 percent of 20-year-olds were enrolled in tertiary education in China, compared to the OECD average of 18 percent, 34 percent, and 39 percent. However, the numbers are still below average for older individuals. For example, only 1 percent of 25–29-year-olds are enrolled in tertiary education, compared to the OECD average of 10 percent (OECD, 2016).

3. Economic Theory

Two theories that relate to an individual's motivation to pursue additional education are Gary Becker's Human Capital Theory and Michael Spence's Job-Market Signaling Model. In this section we will give a brief overview of these two theories and explain how they helped us to formulate our hypothesis.

3.1 Human Capital Theory

"Human capital consists of the knowledge, skills, and health that people invest in and accumulate throughout their lives, enabling them to realize their potential as productive members of society. Investing in people through nutrition, health care, quality education, jobs and skills helps develop human capital" (The World Bank).

One of the most influential economists within Human Capital Theory is Nobel laureate Becker who has published several recognized books within the field (The Nobel Prize, 1993). According to Becker, one of the key components of human capital is education, as it facilitates access to high income jobs and reduces time spent as unemployed, resulting in overall higher income levels (Becker, 1993). Further, his research suggests that investment in education generates marriage market returns since individuals with potential for a high future income or a prestigious job are evaluated as more attractive partners on the marriage market (Becker, 1981).

Becker's research can help us understand individuals' willingness to accumulate education. The fact that taking on more years of schooling leads to an increasing probability of financial stability could encourage males to educate themselves, in order to improve their attractiveness towards potential female partners. This behavior might be spurred as the marriage market competition increases, resulting in males taking additional measures to appear as attractive marriage candidates.

3.2 Job-Market Signaling Model

Another theory that can facilitate the understanding of individuals' motivations is Spence's Job-Market Signaling Model. The model does not discuss actual skills and competences but rather considers how an individual can convey information about their competence level to other parties, under situations with prevailing asymmetric information (see definitions & abbreviations).

Although the connection between the job market and the marriage market may not be obvious, they both share similarities in terms of asymmetric information. On the marriage market, it might be difficult to assess a potential partner's true qualities. Likewise, in most job markets the employer is not 100 percent sure of the skills of a worker at the time of hiring. The fact that it takes time to comprehend an individual's actual skills means that a long-term investment (hiring) is made under uncertainty. However, the individual will have some observable attributes that the less informed party can base its decision on, such as educational attainment. Education signals ability and ambition, therefore, it might serve as an indication and a guarantee for the less informed party that the individual is worth investing in (Spence, 1973).

In the context of the Chinese marriage market, the Job Market Signaling Model suggests that males might feel motivated to invest in education in order to use it as a certificate to signal quality. For example, through obtaining additional education, a male can signal to potential partners that he has high earning potentials (Chang et al., 2011; Oreopoulos and Salvanes, 2011). Further, as it is custom in traditional Chinese marriage arrangements that the wife moves into an accommodation provided by the husband, this might affect the importance of the husband's financial status.

4. Literature review

Chinas skewed sex ratio, its causes and consequences has been thoroughly discussed in previous research. As the existing literature is extensive, it is impossible to present a complete review in this paper. Instead, we have examined a selection of publications that provide relevant insights to our investigation. We have reviewed previous research in four different areas: gender imbalance and education, imbalanced sex ratios and its societal impact, sex ratios and marriage market behavior and finally the relationship between marriage and education.

4.1 Gender imbalance and education

Several studies examine the topic of gender imbalance in relation to education, although the results are somewhat contradicting. Those of interest to this study, investigate how the rising sex ratio imbalance, OCP and son preference have affected educational attainment among males and females. Reviewing this topic provides important insights into how cultural mechanisms together with family planning policies can affect investments in males' education.

In the article "Making Yourself Attractive: Pre-Marital Investments and the Returns to education in the Marriage Market", Canadian economist Lafortune investigated how changes in the sex ratio influenced the pre-marital educational investments among second generation immigrants in the US, born between 1885-1915. Similar to our study, Lafortune applied an instrumental variable

approach where variations in immigration flows between different states in the US were used as instruments for the sex ratio. In particular, she investigated educational attainment among males, and found that in areas with skewed sex ratios, males acquired more formal education. Lafortune found that a one-point standard deviation change in the sex ratio resulted in 0.2 years increase in education for second-generation immigrant males. In summary, her findings suggest that the sex ratio in the surrounding area is of importance when an individual decides whether to undertake additional education or not (Lafortune, 2013).

Gender imbalance and educational attainment is further discussed in "Birth order and the gender gap in educational attainment", by economist Oliveira. She examined both birth order effects on completed years of schooling in China and differences between genders, using the China Health and Retirement Longitudinal Study. Her findings suggests that the OCP, despite exaggerating the skewed sex ratio in China, actually helped reduce the gender gap in educational attainment. Her research also showed that smaller family sizes, as a result of the fertility limits imposed by the OCP, showed a stronger positive effect on human capital investments in females compared to males. As a consequence, a daughter's years of schooling decreased with the number of younger siblings, while a son's schooling increased with the number of younger siblings, hence the OCP led to a reduction in the gender gap in education between the sexes (Oliveira, 2018).

China's prevailing son preference together with perceived lower returns to investments in girls, were explored in the article "Son preference and educational opportunities of children in China—'I wish you were a boy!'". Professor Wang examined the effect of parental son preference on investments in education of boys and girls in China using the 1990 Population Census and ethnographic data. She found that a parental son preference led to lower investments in girls' education in rural China, both in terms of school enrollment and completed years of schooling. The effect was driven mainly by rural parents perceiving lower returns to education for girls, caused by the general gender income gap and traditional marriage customs (Wang, 2005).

4.2 Imbalanced sex ratios and implications for society

Another branch of research has focused on determining the effects of the increasingly skewed sex ratio on society. A common point of interest is the effect of intensified marriage market competition when the number of males relative to females increases. Reviewing this topic highlights the importance of continuous exploration of the field, as existing literature so far suggests that increased marriage market competition has a negative effect on males, and society at large, through e.g., increased crime rates and health problems among males. Linking these findings to the topic of this study can help explain males' motivation to attain additional education as a measure to combat involuntary solitude.

In the article "Too many men, too-short lives: the effect of the male-biased sex ratio on mortality", economics Professors Zhang, Chang and Kan investigated the long-term effect of a male-biased sex ratio on mortality, studying the removal of the marriage ban in Taiwan 1959, using an instrumental variable approach. The study found that males who were more exposed to higher marriage market competition in early adulthood experienced a rise in the probability of cancer, high blood pressure and mortality a few decades later. The study concluded that other countries

experiencing large gender imbalances may suffer corresponding societal costs, stemming from a higher rate of health issues and mortality amongst the male population (Zhang et al., 2020).

A similar point was made by economists Zhang, He and Ma who examined the effect of a skewed sex ratio on males' mental health in China, using the China Family Panel Study dataset from 2012. Similar to our study, the authors instrumented the sex ratio using Ebenstein's (2010) financial penalties under the OCP. The study found that an increasingly biased sex ratio correlated positively with higher depressive symptoms and probability of depression among Chinese men (Zhang et al., 2021).

A further contribution was made by Professors Edlund and Zhang. They investigated the relationship between a skewed sex ratio and the increasing crime rates in China. The authors found that the increasing gender imbalance may account for up to one-seventh of the overall rise in violent and property crime in China during 1988–2004. The authors concluded that the consequences of skewed sex ratios reach beyond males who are unable to find a partner to marry, but also affect society at large (Edlund and Zhang 2013).

4.3 Sex ratio and marriage market behavior

Another effect of the challenges for males to find a partner is the increased incentives to become an attractive marriage candidate. This topic has interested several researchers. Examining these mechanisms can provide insight into how a skewed sex ratio drive male behavior, something which is of interest to our empirical analysis.

Professors Wei, Zhang, and Liu investigated the relationship between differences in sex ratio and housing prices in their article "Home ownership as status competition: Some theory and evidence". They examined house ownership as a key determinant of an individual's marriage market attractiveness in China. When investigating real estate prices in 35 cities in China during 2003-2009, the authors found that housing prices were significantly higher in areas with skewed sex ratios, after controlling for average household wealth, while no such pattern existed for rental homes. Further, their research showed that families having an unmarried son and living in areas with skewed sex ratios, were positively correlated with the probability of the family wanting to own an expensive home. The authors concluded that a higher sex ratio accounted for a significant part of the increase in real estate prices in the areas examined (Wei et al., 2016).

Authors Li, Song, Xu, and Yi examined sex ratios and household investment choices in China. They found evidence that higher sex ratios lead to families with a son having larger marriage expenditures. As a result, in order to appear financially attractive for potential marriage partners, families were more willing to take on risk through investments in high risk / high reward types of assets (Li et al., 2019).

Further, Wei and Zhang examined household savings across China. The authors found that households with a son were more likely to have larger savings in areas with more imbalanced sex ratios, in order to help the son to attract a female partner, and to be able to afford the costly marriage expenditures. Based on their empirical analysis, they concluded that the increased sex

ratio imbalance could explain approximately 60 percent of the increase in household savings during 1990-2007 (Wei and Zhang, 2011b).

In their working paper from 2011, Professors Wei and Zhang investigated the effect of skewed sex ratios on entrepreneurial activities and GDP growth in China. Through applying Ebenstein's (2010) financial penalties, the authors, using an instrumental variable approach, found that a significant part of the GDP growth stemming from private firms could be attributed to an imbalanced sex ratio. Further, the authors found that males in areas with imbalanced sex ratios were more eager to engage in entrepreneurial activities, supplied more labor and had a larger willingness to take (higher paying) uncomfortable and dangerous jobs, all in the hope of acquire additional wealth (Wei and Zhang, 2011a).

4.4 Relationship between marriage and education

A final branch of research investigates another topic closely correlated to this study: the relationship between the marriage market and human capital acquisition. There are typically two subdivisions: one exploring incentives for investments in human capital and the other exploring the probability of marriage (Hu and Sun, 2021). Several scholars have concluded that educated individuals do not only become more attractive on the labor market, but on the marriage market as well, due to increased probability of economic and non-economic returns (Lafortune, 2013; Oreopoulos and Salvanes, 2011, Chang et al., 2011).

In the article "Priceless: The Nonpecuniary Benefits of Schooling", economics Professors Oreopoulos and Salvanes investigated the non-monetary benefits related to education in the US. Their study suggested that educated individuals were more appealing on the marriage market, both due to increased probability of higher earnings and occupational prestige, but also through improved cognitive skills, such as critical thinking and social skills acquired in school (Oreopoulos and Salvanes, 2011).

A similar point was made by Chang, Wang, Shackelford, and Buss. The authors examined the development of mating preferences in Shanghai comparing survey responses from individuals in 1982 to 2008. One important finding was that a college degree increased an individual's attractiveness for both males and females, although the effect was slightly larger for males. Further, the authors found these preferences stronger among the 2008 sample compared to the 1982 sample, indicating that this preference had become more significant over time. Similarly, high earnings capacity and good financial outlooks were found to be important for an individual's attractiveness. This preference had also risen in parallel with education, leading to the authors suggesting that these may correlate. Finally, the authors concluded that females' value financial resources in a potential partner more than males (Chang et al., 2011).

The existing research on the success of highly educated individuals is, however, not univocal. In the article "Too educated to be married? An investigation into the relationship between education and marriage in China", professors Hu and Sun investigated the relationship between education and marriage market success in China. They found that individuals with a college degree were less likely to be married for two reasons. Firstly, highly educated individuals had higher expectations

on their potential partners. Secondly, as for other groups, increasing marriage costs, driven by accelerating house prices, was found as a discouraging factor (Hu and Sun, 2021).

5. Contribution

5.1 Research gap, importance of the topic and contribution to literature

From previous sections it can be concluded that China's skewed sex ratio could have severe consequences for males' marriage outlooks as well as their overall health and well-being. While the literature present rich evidence that a skewed sex ratio causes behavioral changes among males, motivated by intense marriage market competition, little research investigates the direct link between increased marriage market competition and investments in education. Existing literature presents contradictive findings regarding the implications of increased education on marriage prospects. Some studies have concluded that education improves an individual's attractiveness to potential partners while other studies have found that higher education results in lower probability of marriage. In this paper, we have attempted to contribute to shed further light on this topic, by investigating the relationship between a skewed sex ratio and Chinese males' motivation to use education as a tool to signal attractiveness to potential partners. To our knowledge, this potentially causal relationship has not been examined before.

Regarding the contribution to existing literature, Lafortune (2013) is the only paper, to the best of our knowledge, to examine the relationship between an imbalanced sex ratio and male's attempt to increase their attractiveness on the marriage market through education. However, the scope of that study is second generation immigrants in the US, during the early 1900's. Hence, we contribute to existing research by extending the knowledge of the relationship to a Chinese 21st century context, a geographical and temporal research setting which currently is unexplored. Our analysis is also demographically broader, since we are analyzing a data sample representing the whole Chinese population of marriageable age. Furthermore, the data used in our thesis is conducted during the 21st century, thus enabling us to draw conclusions about the current Chinese context, which might be helpful for continuous investigations and policy makers interested in this topic.

The conclusions of this study may further provide insights into parents' perceived return on investments in their children. Investigating this topic can thus be an important contribution from a gender equality point of view, as it can provide guidance to policy makers on factors affecting educational inequality between genders. In addition, as sex ratios vary among provinces, our study could provide insights into the underlying causes of regional differences in males educational attainment, which can be helpful in trying to reduce the educational gap across China.

5.2 Hypothesis and research question

Previous literature points out that a skewed sex ratio is causing males to change their behavior, in order to improve their marriage market competitiveness. As males' financial status has been shown to increase their attractiveness from the female point of view, these behavioral changes have in previous literature been motivated by a desire for increased wealth (Li et al., 2019; Wei and Zhang,

2011a; Chang et al., 2011; Becker, 1981). Furthermore, Human Capital Theory stipulates that education facilitates access to higher paying jobs and increased lifetime earnings. Moreover, the Job Market Signaling Model provides an explanation to how education can be used to signal high-earning potentials in situations where asymmetric information prevails. Hence, from the discussion above, the following hypothesis and research question has been formulated:

Hypothesis: A skewed sex ratio will increase males' motivation to invest in education in order to signal their attractiveness to potential (female) partners.

Research question: How does a skewed sex ratio impact males' investments in education?

6. Data

This section presents an overview of the data used in this study and definitions of the main variables in our analysis.

In order to evaluate the effect of an imbalanced sex ratio on males' length of education, we have combined several datasets. The first one is China Family Panel Studies (CFPS) from 2010. This dataset is a national, cross-country survey conducted biannually from 2010 and onwards, by the Institute of Social Science Survey (ISSS) of Peking University in China. The aim of the survey is to collect information about economic, non-economic and social aspects of the Chinese population. The survey is based on personal interviews and consists of individual, household, and community level responses from China's 31 provinces in both rural and urban China (see table (12) for list of China's provinces). The interviews were conducted on approximately 30 000 individuals with a 79 percent overall response rate. On specific questions, the data quality is much lower, resulting in a large proportion of "missing values" in the dataset (CFPS, 2022).

Our outcome variable, "Years of education", is based on data from the 2010 CFPS. This variable captures an individual's completed years of education in 2010, for the age group 16-35. This is our primary group of interest for two reasons. Firstly, education is compulsory during the first nine years of schooling which approximately corresponds to the age of 15. Individuals obtaining additional education can, therefore, be observed first after that age. Secondly, the 16-35 age group covers the majority of the people marrying in China, as the average marriage age is 24 which falls well within our sample range (Zhang et al., 2021)[6].

Our treatment variable, "Sex ratio for cohort 16-35", is compiled using data from the China Population Censuses from 1990 and 2000. The population censuses have been organized by the Chinese government since the 1950's and are mainly conducted through the National Bureau of Statistics of China and other statistical bureaus (Zhang and Zhao, 2006). The censuses capture demographic, economic and social information about the Chinese population (IPUMS).

[6] Similar age cohorts have been used as the prime marriage age in previous studies. Zhang, He & Ma (2020) uses age cohort 17-26, while Cameron, Meng & Zhang (2019) look at ages 18-27. Edlund, Li, Yi & Zhang (2013) uses age groups 16-25. Due to our topic of analysis, years of education, where a doctoral degree can be obtained at the age of 27 at the earliest, we have chosen to expand our sample, compared to other studies.

Using the population censuses, we are able to calculate the ratio of males per 100 females for ages 16-35 in 2010 (aged 6-15 in 1990 and 2000) in each province[7]. As the marriage market is mainly "local", it is reasonable to assume that individuals are affected by the sex ratio in their province of residence (Zhang et al., 2021). Due to the Chinese hukou system, individuals migrating abroad or moving to other provinces prior to the age of 15, lose certain public benefits like access to public schools, which is why we assume that the "migration effect" is insignificant (Cui et al., 2019). Thus, we can postulate that the sex ratio during an individual's youth remains that of the birth province. We hypothesize that the youth sex ratio, and hence, the individual's future marriage market competitive situation, will affect the individual's inclination to acquire additional, non-compulsory, years of education, following the age of 15. Therefore, an individual's youth sex ratio at their province of birth is a valid indication of the marriage market pressure they experience growing up[8].

It is likely that individuals look beyond their own age group in the mating process, meaning that they are also affected by the sex ratio in adjacent age groups. To account for this effect, our sex ratio variable is constructed as a five-year age mean in each province of birth for individuals aged 16-35. Through this method, we obtain four different sex ratio cohorts: 16-20, 21-25, 26-30 and 31-35[9]. In addition, we have compiled an additional sex ratio variable where we have excluded outlier values, to make sure that our calculated effects are not entirely driven by outliers in the sex ratio (see figure (3) & (4) in appendix).

Through combining the 2010 CFPS and the sex ratio data, we can create our main variable of interest; an interaction variable for the average sex ratio and the gender effect for males. Control variables at the individual and household levels are derived from the 2010 CFPS survey. County level characteristics are obtained from a restricted extension of the CFPS, which contains county level information (CFPS, 2022). By combining the 2010 CFPS survey together with the 1990 and 2000 Population Censuses, we end up with a main analysis sample consisting of 8579 individuals, aged 16-35, from 30 provinces[10].

7. Empirical Method

This section begins with a presentation of our empirical models and motivation for our chosen control variables. Further, we elaborate on potential weaknesses of the Ordinary Least Squares (OLS) model which are controlled for through an Instrumental Variable (IV) model. Finally, we

[7] We have calculated the sex ratio for the age group 6-15 in year 1990 and matched it to individuals aged 26-35 in 2010. Similarly, we have calculated the sex ratio for the age group 6-15 in year 2000 and matched it to individuals aged 16-25 in 2010.

[8] Although it is possible that individuals migrate to another province after completing compulsory schooling, which suggests that they could be faced with a different marriage market competition, we assume that this effect is minor as individuals' current province of residence and province at birth have a 97.36 percent correlation in our dataset.

[9] Other studies (Zhang, He & Ma, 2021; Zhang & Wei, 2011) calculate an average sex ratio for the whole sample age. However, due to the limited number of provinces in China, the five-year averages also serve the purpose of increasing the variation within our sample.

[10] One province did not contain any respondents from our age group and is therefore not present in our main sample.

discuss the required IV assumptions and introduce our chosen methods to test for instrument validity.

7.1 The baseline model: OLS regression

To examine the effect of a skewed sex ratio on males' years of education, we start by estimating the linear relationship between the two, through an OLS model.

Our baseline model is the following:

$$Y_{itpc} = \beta_0 Male_i + \beta_1 Sex\_Ratio_{tp} + \delta X_{itpc} + \varepsilon_{itpc} \quad (1)$$

$$+ \beta_2 Male_i * Sex\_Ratio_{tp} \quad (2)$$

$$+ \alpha_p \quad (3)$$

Where Y_{itpc} represents the outcome variable, completed years of education of respondent i, from cohort t, in province p, and county c. Male is an indicator variable for male respondents, of individual i. The treatment variable Sex_Ratio_{tp} refers to the sex ratio of cohort t in province p. X_{itpc} represents our set of demographic and socio-economic characteristics at the individual level i, cohort t, province p, and county level c (1). The interaction variable $Male_i * Sex\_Ratio_{tp}$ represents the effect of the sex ratio for male respondent i, from cohort t in province p. (2). Further, we add a control variable for province specific fixed effects, α_p, in order to control for between province variations (3). In line with the reasoning proposed by Bertrand, Duflo and Mullainathan (2004) we cluster our standard errors at the level of treatment, cohort t and province of p.

In addition, we add a regression with trimmed (excluding outliers) sex ratio values, in order to rule out that our results are not entirely driven by outliers (see figure (3) & (4) in appendix)[11]. The above model is thus adjusted accordingly:

$$Y_{itpc} = \alpha_p + \beta_0 Male_i + \beta_1 Sex\_Ratio\_trim_{tp} + \beta_2 Male_i * Sex\_Ratio\_trim_{tp} + \delta X_{itpc} + \varepsilon_{itpc} \quad (4)$$

7.1.1 Splitting dataset by hukou

As hukou status is an important determinant of an individual's access to the social welfare system, we further want to examine differences between agricultural and non-agricultural hukou individuals. Therefore, in addition to our main dataset, we created two new datasets, splitting our original data by hukou status.

7.1.2 Control variables

The chosen control variables in this model (see below) are all believed to have an impact on our dependent variable "Years of education". Including these variables helps increase the precision of

[11] Reducing the distribution of the sex ratio from 92.02-124.92 to 102.82-119.50.

our estimates and will most likely result in higher explanatory value of the model as a whole. Excluding variables affecting our dependent variable would result in the model attributing the effect of those omitted variables to the included ones, resulting in biased estimates. If these missing variables happen to be correlated with the sex ratio, we would detect a relationship between educational attainment and the sex ratio, although it might not exist. By including the control variables, we try to isolate the sex ratio's exogenous effect on the dependent variable and, thereby, establish a reliable estimation of the potential causal effect sought for in our research question.

The control variables are set to capture demographic and socio-economic characteristics of the individuals in our sample. The gender of the respondent is included to control for differences in educational attainment among males and females, as males constitute our main point of interest. A variable for area of residence is included to account for the differences between rural and urban residency, as urban areas are believed to have better access to schools compared to rural ones (Ayoroa et al., 2010). Similar to the previous control, the individual's childhood hukou status is included to control for access to education through distinguishing between agricultural and non-agricultural families (see section 2.3 for explanation of hukou). As mentioned in by Chen & Silverstein (2000), rural families working with agriculture are more dependent on their sons as labor source and old-age support which might lead to different educational investments in their child(ren) depending on gender.

A variable for the individual's mother's level of education is included to consider the effect of parent's education on their children's educational attainment, as previous research has found a positive correlation between the two (Harding et al., 2015)[12]. Further, another determinant of an individual's educational attainment is the family income, which is why log of family income is controlled for in this model (Blanden and Gregg, 2004). Lastly, log of county level GDP per capita is included to cater for the effect of regional wealth differences among the population, which might affect educational attainment, although this measure is an imperfect proxy (Kapoor and Debroy 2019)[13]. Finally, it is worth noting that a majority (> 80 percent) of the respondents in our dataset possess an agricultural hukou status while the other variables have a more dispersed distribution of values.

7.1.3 Weaknesses of the OLS model

There are several potential problems associated with using an OLS regression model to detect the existence of a causal relationship between a sex ratio imbalance and years of education for males. Although we try to control for the determinants of years of education, there are most likely other predictors that are not part of our model. As discussed in section 7.1.2, omitted variables result in biased estimates. One commonly mentioned example is China's prevailing preference for sons (Zhang et al., 2021). It is possible that a son preference influences both the sex ratio and educational attainment among males, as parents with this preference might be more prone to engage in sex selective activities as well as invest more in their son's education compared to other

[12] We chose to exclude a control for father's level of education as it would reduce our sample size significantly. Additionally, including this control did not affect our results substantially.

[13] We are using the logarithm of family income and GDP per capita in order to linearize the relationship with years of education.

parents. In addition, due to the large number of "missing values" in the CFPS dataset, other potential determinants of educational attainment were not possible to include in our model (see section 9.2.1 for further elaboration). Since these variables remain uncontrolled for, the model might be further biased as a consequence.

A second problem with the OLS model is potential measurement errors in the sex ratio reporting. Underreporting of girls has been detected in China, which could result in the sex ratio obtained from the Population Censuses not reflecting the actual sex ratio in China (Goodkind, 2011) . Measurement errors in the treatment variable cause endogeneity problems and result in biased estimates. However, underreporting of girls is unusual after the age of six due to the household registration system (hukou), being a requirement for entering compulsory schooling at the age of six (Zhang et al., 2021). Our method of calculating the sex ratio for individuals aged 6-15, allows us to mitigate this issue.

7.2 Controlling for weaknesses in the OLS model: adopting an IV approach

To address the potential endogeneity issues, we are adopting an IV approach[14]. According to this model, we try to isolate the effect of the sex ratio on educational attainment, through examining determinants of the sex ratio that are assumed to be uncorrelated with the dependent variable and the error term. This permits us to draw conclusions about the potential relationship between the sex ratio and years of education.

Previous research attributes the OCP as one of the main causes of the imbalanced sex ratio in China (Chen et al., 2013; Sen, 1990). Ebenstein (2010) introduced a method, using regional differences in financial penalties for violating the imposed fertility limits, as instruments for the sex ratio. This was built on the hypothesis that stricter financial penalties influence parents to engage more in sex-selective activities instead of having to pay fines for additional children. We are using two financial penalty types proposed by Ebenstein (2010): the fine rate for violating the OCP expressed as multiples of annual income, and the additional premium fines for higher order births[15]. In the model, the average fine rates and premiums for years 1985-1994 are used as instruments for the sex ratio of individuals aged 16-25 in the 2010 CFPS sample, while average fine rates and premiums for years 1979-1984 are used as instrument for the sex ratio for individuals aged 26-35 in the 2010 CFPS [16].

[14] In addition, the IV model serves as a test of whether our findings from the OLS regressions are consistent after a change of method, which could provide additional support for our results.

[15] For example, if a family under a strict one-child limit chose to have three children, they could be faced with larger penalties compared to if they "only" had two (Ebenstein, 2010).

[16] These years correspond to the years in which the individuals in the cohorts were born. Cohorts 16-20 and 21-25 in 2010 CFPS were born 1985-1994 while cohorts 26-30 and 31-35 were born between 1975-1984. However, as the OCP and financial penalties were implemented in 1979, we use the years 1979 and 1984 as an imperfect approximation for the age group 26-35. We use the average fine rate and premiums in order to account for the fact that the decision to violate the fertility limits depends not only on the prevailing fine rate and premium at the year of giving birth, but of those prevailing previous years as well. This method is inspired by previous research from scholars such as Zhang He and Ma (2021), Wei and Zhang (2011a). We use one single average for age group 6-15 in 1990 and another for age group 6-15 in 2000 as the penalties were fairly constant over time.

Financial penalty data was retrieved from Ebenstein at the Hebrew University of Jerusalem. This data has been used in previous research, i.e. Zhang, He and Ma (2020), Wei and Zhang (2011a) and Ebenstein (2010); all examining the sex ratio in China (Hebrew University of Jerusalem).

According to this, we estimate the first-stage regressions as the following:

$$Sex\_Ratio_{tp} = \alpha_p + \delta_0 Fine\_Rate_{tp} + \delta_1 Premium_{tp} + +\pi X_{itpc} + \epsilon_{itpc} \quad (1)$$

$$Sex\_Ratio\_trim_{tp} = \alpha_p + \delta_0 Fine\_Rate_{tp} + \delta_1 Premium_{tp} + +\pi X_{itpc} + \epsilon_{itpc} \quad (2)$$

Where Sex_Ratio_{tp} is the average sex ratio for cohort t in province p. $Fine\_Rate_{tp}$ is the fine rate for cohort t in province p. $Premium_{tp}$ is a binary variable taking on the value of 1 if a premium for additional violation of OCP existed for cohort t in province p. X_{itpc} represents our set of demographic and socio-economic characteristics at the individual i, cohort t, province p, and county level c. Furthermore, we control for province fixed effects α_p. Standard errors are clustered at the province and cohort level. Similar to the OLS regression, we are conducting an additional regression testing based on the trimmed sex ratio.

7.2.1 IV Model assumptions

In order to conduct an IV analysis, two assumptions have to hold: relevance and exogeneity.

1) $Relevance: Cov(z, x) \neq 0$

An instrument is relevant if sufficient covariance exists between the independent variable, in our case the treatment variable "Sex Ratio for cohort 16-35", and the instrument "Fine Rate" or "Premium", otherwise the instrument is considered "weak". A general rule of thumb is that instruments are not weak if the first-stage F-statistic exceeds the value of 10 (Lee et al., 2020). However, Andrews, Stock and Sun (2018) have proposed that a more appropriate threshold lies around the value of 50. In addition to Andrews, Stock and Sun, economists Stock and Yogo (2005) have compiled a set of critical values, to which the F-statistic can be compared to under the null hypothesis that the instruments are weak. If the F-statistic is larger than those critical values, the null hypothesis of weak instruments can be rejected. To test for weak instruments in our model, we obtain the Kleibergen–Paap rk F-statistic and compare it to Stock and Yogo's critical values[17].

2) $Exogeneity: Cov(z, u) = 0$

The second assumption of an instrumental variable approach is exogeneity. An instrument is exogenous if it is uncorrelated with the error term and, therefore, has no effect on the dependent variable. It is plausible that the financial penalties under the OCP solely correlates with the sex ratio, since these specific penalties were applied to individuals violating the imposed fertility limits, which raises doubt if they have an impact on the other variables in the model. Exogeneity can be tested when a model has more instruments than treatment variables (an "overidentified" model), which is the case in our investigation. We will perform Hansen's J-statistic calculation in order to

[17] The Kleibergen–Paap rk F-statistic is an applicable F-statistic when standard errors are clustered.

test the joint null hypothesis that the instruments are exogenous, a test that has been applied by previous scholars using the financial penalties, proposed by Ebenstein (2010), as instruments for the sex ratio in China, (Zhang et al., 2021; Wei and Zhang, 2011a).

If the Kleibergen-Paap rk Wald F-statistic is above Stock and Yogo's critical values and the p-value from the Hansen's J Statistic cannot reject the null hypothesis of exogenous instruments, it is an indication of valid instruments. The results of the Kleibergen–Paap rk and Hansen's J statistics are found in our output tables.

7.3 Descriptive statistic

Table (2) presents our descriptive statistics. The average years of education for individuals aged 16-35 in 2010 is 9.6 years, with a standard deviation of 3.9 years. The sex ratio averages range between 92.0 and 124.9, which suggest that the sex ratio vary considerably between provinces. Also, the mean value of 108.2 is higher than the expected or "natural" sex ratio of 102-107 males per 100 females[18].

Table 2 Descriptive Statistics

	N	Mean	Median	Std. Dev.	Min	Max	Dataset
Dependent variable							
Completed years of education in 2010	8579	9.605	9.000	3.896	0	22	CFPS
Independent variables							
Sex ratio for cohort 16-35	8579	108.157	108.313	4.595	92.022	124.92	Census
Sex ratio for cohort 16-35, trimmed	8133	108.545	108.566	3.989	102.819	119.5	Census
Male[1]	8579	.484	0.000	.5	0	1	CFPS
Sex ratio for cohort 16-35 x male	8579	52.418	0.000	54.191	0	124.92	CFPS/Census
Sex ratio for cohort 16-35 x male, trimmed	8133	52.423	0.000	54.371	0	119.5	CFPS/Census
Urban residence [2]	8579	.472	0.000	.499	0	1	CFPS
Hukou status at age 3 [3]	8579	1.392	1.000	1.182	1	79	CFPS
Mother's level of education[4]	8579	2.016	2.000	1.077	1	8	CFPS
Log. of family income	8579	10.231	10.251	.924	5.298	14.529	CFPS
Log. of GDP per capita	8579	10.102	9.881	1.058	8.062	12.676	CFPS
Instruments							
Fine rate	8579	1.301	1.214	.378	.453	2.388	Ebenstein
Premium[5]	8579	.572	0.667	.392	0	1	Ebenstein

[1] Male = 1; Female = 0
[2] Urban = 1; Rural = 0
[3] Agricultural = 1; Non-agricultural = 3; No registration = 5; Non-Chinese Registration = 79
[4] Illiterate = 1; Primary educ = 2; Junior educ = 3; Senior educ = 4; College = 5; Bachelor = 6; Doctor = 8
[5] Premium = 1; No premium = 0

Source: CFPS (2022); IPUMS; Hebrew University of Jerusalem

Table (3) shows that the average years of schooling is higher for males in our sample compared to females, but also that the distribution of educational attainment varies more amongst females (see distribution in figures (7) & (8) in appendix). In absolute numbers, the highest years of education in the female sample is larger than the male sample maximum (22 years compared to 19 years).

[18] The expected sex ratio at birth in the absence of interference lies between 103 to 107 boys per 100 girls (Ritchie and Roser, 2019).

Table 3 Summary Years of Education in 2010

	N	Mean	Median	Std. Dev.	Min	Max
Males	4155	9.867	9.000	3.592	0	19
Females	4424	9.36	9.000	4.146	0	22

Source: CFPS (2022)

8. Empirical results

In this section we present the results from our econometric models investigating the sex ratio effect on males completed years of schooling. We start by presenting our results from the OLS regressions in table (4). In column (1), we report the baseline model with controls, excluding the interaction term "Sex Ratio for cohort 16-35 x Male". Continuing, in column (2) we add the interaction term, in column (3) we control for province fixed effects and finally, in column (4), we exclude extreme values in the sex ratio variable (see figure (3) & (4) appendix). In addition to our main regression, we include an additional table (table (5)) where the sample is split by hukou status. Further, we continue this section by reporting the first and second stage of our IV model, where the datasets are split by gender. In column (1) of the first and second stage model we regress the regular sex ratio while column (2) reports the trimmed version.

8.1 Main results

Table (4) reports our main results from the OLS regression, where we regress the sex ratio on completed years of education.

Table 4 OLS Regression of Youth Sex Ratio on Years of Education in 2010

	(1) **Years of education**	**(2)** **Years of education**	**(3)** **Years of education**	**(4)** **Years of education**
Male	0.323***	5.463***	5.193***	4.967**
	(0.0903)	(1.945)	(1.892)	(2.282)
Sex ratio for cohort 16-35	0.0296**	0.0543***	0.0565***	0.0626***
	(0.0117)	(0.0153)	(0.0153)	(0.0180)
Sex ratio for cohort 16-35 x male		-0.0474***	-0.0448**	-0.0426**
		(0.0179)	(0.0174)	(0.0209)
Urban residence	1.210***	1.210***	1.293***	1.274***
	(0.110)	(0.110)	(0.112)	(0.115)
Hukou status at age 3	0.247**	0.244**	0.277**	0.277**
	(0.108)	(0.108)	(0.129)	(0.133)
Mother's level of education	0.969***	0.971***	0.889***	0.885***
	(0.0560)	(0.0559)	(0.0546)	(0.0559)
Log. of family income	0.660***	0.656***	0.652***	0.686***
	(0.0595)	(0.0595)	(0.0598)	(0.0624)
Log. of GDP per capita	0.692***	0.693***	0.525***	0.477***
	(0.0573)	(0.0574)	(0.0593)	(0.0627)
Constant	-10.36***	-13.01***	-11.32***	-11.77***
	(1.678)	(1.968)	(2.181)	(2.521)
Observations	8,579	8,579	8,579	8,133
R-squared	0.315	0.315	0.348	0.352
Province FE	NO	NO	YES	YES

Notes: The dependent variable refers to completed years of schooling in 2010. The sex ratio is calculated as a five-year mean value for ages 16-35 in each province. The sex ratio for the age cohort 16-25 in 2010 is inferred from age cohort 6-15 in the 2000 population census. The sex ratio for the age cohort 26-35 in 2010 is inferred from age cohort 6-15 in the 1990 population census. In column (1), interaction term between sex ratio and gender is not included. In column (4), average sex ratio excludes outliers. Log. of GDP per capita refers to county GDP per capita. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS

Despite the fact that the control variables are not our main area of interest, several of them are worth commenting on. One is the "Male" variable, which increases from 0.32 in column (1) to 5.46 in column (2). Without controlling for the interaction variable "Sex ratio for cohort 16-35 x male", males attain 0.32 more years of education than females. However, when including the interaction variable in column (2), the coefficient for "Male" reports the difference in intercept between male and females in areas where the sex ratio is 0. As this is an impossible occurrence in reality, we do not interpret this coefficient further.

The variable "Urban residence" remains significant at the 0.01 level across all four regressions. It reports the highest coefficients in column (3) of about 1.29 years of additional schooling for urban residents compared to rural inhabitants. Furthermore, the variable "Hukou status at age 3" reports a positive correlation with the dependent variable across all regressions and a statistical significance at the 0.05 level. This indicates that non-agricultural hukou status individuals attain more years of education, as non-agricultural hukou status is represented by higher values in the dataset (see descriptive statistics in table (2)). The variable "Mother's level of education" is statistically significant at the 0.01 level for all regressions and correlates positively with the variable "Years of education". Column (2) shows that a one level increase in the mother's level of education increases an individual's educational attainment with 0.97 years. The variable "Log. of family income" also exhibits a positive correlation with the dependent variable and is significant at the 0.01 level across all regressions. All columns report that increases in the family income leads to increased years of schooling. In addition, the variable "Log. of GDP per capita" reports statistical significance at the 0.01 level and a positive correlation with the dependent variable for all four regressions. All control variables have positive beta coefficients, implying that they correlate positively with the outcome variable "Years of schooling".

The treatment variable "Sex ratio for cohort 16-35" has a positive coefficient, meaning that as the gender imbalance increases, the overall educational attainment does as well. Across all four columns, the variable is significant at the 0.05 level in column (1) and at the 0.01 level in column (2), (3) and (4). The key explanatory variable, "Sex ratio for cohort 16-35 x Male" has a negative beta coefficient across all columns. This suggests that when the sex ratio increases, males attain less education compared to females (although the overall educational level increases). In column (2), a one unit increase in the "Sex ratio for cohort 16-35" (e.g., from 107 to 108 males per 100 females) corresponds to 0.05 less years of education for males. This variable is statistically significant at the 0.01 level in column (2), and at the 0.05 level in column (3) and (4).

The R-squared value ranges between 0.32 in column (1) and (2) to 0.35 in column (4), meaning that the model's explanatory value increases as we control for province fixed effects and apply the trimmed sex ratio.

8.1.1 Separating sample by hukou status

Table (5) reports the results from the OLS regression when the original dataset is split by hukou status.

Table 5 OLS Regression of Youth Sex Ratio on Years of Education in 2010, non-agricultural hukou sample

	Non-agricultural sample		**Agricultural sample**	
	(1) **Years of education**	**(2)** **Years of education**	**(3)** **Years of education**	**(4)** **Years of education**
Male	2.101	2.502	5.881***	6.101**
	(3.630)	(3.711)	(2.185)	(2.752)
Sex ratio for cohort 16-35	-0.0748**	-0.128***	0.0828***	0.100***
	(0.0319)	(0.0360)	(0.0163)	(0.0208)
Sex ratio for cohort 16-35 x Male	-0.0217	-0.0256	-0.0500**	-0.0517**
	(0.0333)	(0.0340)	(0.0201)	(0.0252)
Urban residence	3.271***	3.112***	1.074***	1.057***
	(0.390)	(0.382)	(0.112)	(0.116)
Mother's level of education	0.594***	0.594***	0.860***	0.860***
	(0.0851)	(0.0874)	(0.0563)	(0.0573)
Log. of family income	0.736***	0.731***	0.568***	0.608***
	(0.114)	(0.116)	(0.0661)	(0.0694)
Log. of GDP per capita	0.473***	0.466***	0.550***	0.489***
	(0.124)	(0.127)	(0.0657)	(0.0689)
Constant	3.285	9.428*	-13.90***	-15.59***
	(4.662)	(4.972)	(2.653)	(3.078)
Observations	1,560	1,506	7,019	6,627
R-squared	0.346	0.354	0.280	0.283
Province FE	YES	YES	YES	YES

Notes: Regressions are identical to regression (4) but column (1) & (2) included only non- agricultural hukou registered individuals, while column (3) & (4) includes only agricultural hukou registered individuals. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS

When splitting the sample by hukou status of the individuals at their age of three, our findings are somewhat different. In the non-agricultural sample, "Years of education" correlates negatively with the "Sex ratio for cohort 16-35" compared to the agricultural sample where the correlation is positive. This means that when the sex ratio increases, non-agricultural individuals attain *less* education while agricultural individuals attain *more*. These findings are consistent and statistically significant (although at varying levels) across all regressions. The interaction variable "Sex ratio for cohort 16-35 x Male" continues to report negative coefficients in both samples, although they are insignificant in the non-agricultural sample. The control variables continue to report positive and statistically significant coefficients, similar to table (4).

8.2 IV regression

In this section we report the results from the IV regressions. We examined two samples, one containing male respondents and the other female respondents. In both the first- and second stage, column (1) and (3) examine the regular sex ratio while column (2) and (4) analyze the trimmed sex ratio variable.

Table 6 First Stage Regression of Financial Penalties on Youth Sex Ratio

	Males		**Females**	
	(1) **Sex ratio for cohort 16-35**	**(2)** **Sex ratio for cohort 16-35**	**(3)** **Sex ratio for cohort 16-35**	**(4)** **Sex ratio for cohort 16-35**
Fine rate	5.506***	2.468***	5.091***	2.440***
	(0.819)	(0.512)	(0.792)	(0.520)
Premium	3.725***	4.356***	3.772***	4.326***
	(0.557)	(0.531)	(0.504)	(0.472)
Urban residence	-0.122	-0.0821	-0.156	-0.0976
	(0.156)	(0.127)	(0.143)	(0.132)
Hukou status at age 3	0.0513	0.0288	0.0186	-0.0213
	(0.0843)	(0.0691)	(0.0494)	(0.0335)
Mother's level of education	0.346***	0.246***	0.269***	0.264***
	(0.0731)	(0.0649)	(0.0752)	(0.0607)
Log. of family income	-0.353***	-0.255***	-0.0640	-0.0857
	(0.0800)	(0.0661)	(0.0774)	(0.0650)
Log. of GDP per capita	0.0772	-0.221***	-0.105	-0.316***
	(0.142)	(0.0807)	(0.0968)	(0.0851)
Constant	100.9***	108.0***	101.9***	108.2***
	(2.447)	(1.555)	(1.841)	(1.520)
Observations	4,155	3,924	4,424	4,209
R-squared	0.444	0.491	0.417	0.491
Kleibergen-Paap rk Wald F statistic	72.423	75.876	77.129	77.757
Province FE	YES	YES	YES	YES

Notes: The dependent variable sex ratio is calculated as a five-year mean value for ages 16-35 in each province. Column (1) & (2) includes male individuals, while column (3) & (4) includes only female individuals. The sex ratio for the age cohort 16-25 in 2010 is inferred from age cohort 6-15 in the 2000 population census. The sex ratio for the age cohort 26-35 in 2010 is inferred from age cohort 6-15 in the 1990 population census. In column (2) & (4), average sex ratio excludes outliers. Log. of GDP per capita refers to county GDP per capita. Average fine rate and premium for years 1985-1994 are used as instrument for sex ratio for age 6-15 in 2000 (16-25 in 2010). Average fine rate and premium for years 1979-1984 are used as instruments for sex ratio for age group 6-15 in 1990 (26-35 in 2010). For the Kleibergen-Paap rk Wald F test for weak instruments, the Stock and Yogo critical values are: 10% maximal IV relative bias 19.93; 15% maximal IV relative bias 11.59; 20% maximal IV relative bias 8.75; 25 % maximal IV size 7.25. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS; Hebrew University of Jerusalem

In table (6), the first stage regressions are reported. "Sex ratio for cohort 16-35" is regressed on the instruments "Fine rate" and "Premium". Both instruments show statistical significance at the 0.01 level and have positive coefficients across all regressions. This indicates that stricter financial penalties are associated with higher gender imbalance. The instrument "Fine rate" reports that every unit increase in the fine rate (in terms of multiples of annual income) is associated with a 5.51 increase in the province sex ratio for the male sample in column (1), and a 5.09 increase in the female sample in column (3). The existence of additional fines for higher order births, "Premium", is associated with a 3.73 increase in the province sex ratio for the male sample in column (1) and a 3.77 increase in the female sample in column (3). When applying the trimmed sex ratio in column (2) and (4), the "Premium" variable reports higher coefficients of 4.36 in column (2) and 4.33 in column (4) compared to "Fine rate" coefficients of 2.47 in column (2) and 2.44 in column (4).

Furthermore, the Kleibergen-Paap rk Wald F-statistic test for weak instruments is above the Stock and Yogo critical values at the 10 percent level, suggesting that the instruments are strong. The R-squared reports values of between 0.42 - 0.49 in columns (1) - (4).

Table 7 Second Stage Regression of Youth Sex Ratio on Years of Education in 2010

	Males		**Females**	
	(1) **Years of education**	**(2)** **Years of education**	**(3)** **Years of education**	**(4)** **Years of education**
Sex ratio for cohort 16-35	0.0410*	0.0778**	0.170***	0.228***
	(0.0244)	(0.0342)	(0.0299)	(0.0390)
Urban residence	0.977***	1.004***	1.697***	1.637***
	(0.156)	(0.160)	(0.154)	(0.160)
Hukou status at age 3	0.462***	0.460***	0.181	0.196
	(0.0916)	(0.0946)	(0.118)	(0.128)
Mother's level of education	0.670***	0.651***	0.967***	0.948***
	(0.0626)	(0.0662)	(0.0736)	(0.0761)
Log. of family income	0.634***	0.676***	0.667***	0.702***
	(0.0724)	(0.0747)	(0.0829)	(0.0867)
Log. of GDP per capita	0.488***	0.479***	0.588***	0.587***
	(0.0853)	(0.0899)	(0.0837)	(0.0963)
Constant	-7.736**	-12.39***	-24.42***	-31.03***
	(3.039)	(4.153)	(3.712)	(4.729)
Observations	4,155	3,924	4,424	4,209
R-squared	0.305	0.306	0.388	0.390
Hansen J statistic (overidentification test)	0.0593	0.1003	0.1848	0.7959
Province FE	YES	YES	YES	YES

Notes: The dependent variable is completed years of schooling in 2010. Column (1) & (2) includes male individuals, while column (3) & (4) includes only female individuals. In column (2) & (4), average sex ratio excludes outliers. Log. of GDP per capita refers to county GDP per capita. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS; Hebrew University of Jerusalem

The results of the second stage are presented in table (7). The coefficients of the "Sex ratio for cohort 16-35" are significant at the 0.01 level for the female sample and at the 0.1 level in column (1), and 0.05 level in column (2) for the male sample. In all regressions, the variable correlates positively with the dependent variable "Years of education" meaning that the years of education increases for both genders in areas where the sex ratio imbalance increases.

The variables "Urban residence", "Mother's level of education", "Log. of family income" and "Log. of GDP per capita" correlate positively with the dependent variable and are significant at the 0.01 level across all regression in both samples. However, the variable "Hukou status at age 3" is significant at the 0.01 level in both regressions in the male sample but insignificant in the female sample.

Column (1) has a p-value in Hansen's J statistic of 0.06 while column (2), (3) and (4) have p-values of 0.10, 0.18 and 0.80 respectively. This means that the assumption of exogenous instruments cannot be rejected at the 0.05 significance level in all four regressions. As 0.05 is typically considered the threshold for statistical significance (Di Leo and Sardanelli, 2020), both instruments can be considered exogenous and relevant, which suggests that they are valid.

9. Discussion

This section begins with a revision of the research question and hypothesis in comparison to our empirical results. We continue by discussing the economic implications of our findings and how they relate to existing literature. This section carries on with a discussion of the limitations of the study in terms of data and methods. Lastly, we point out what is left to be explored by future researchers.

9.1 Discussion of results

Based on previous literature, Human Capital Theory and the Job-Market Signaling Model we formulated the following hypothesis: *A skewed sex ratio will increase males' motivation to invest in education in order to signal their attractiveness to potential (female) partners.*

Consequently, we intended to answer: *How does a skewed sex ratio impact males' investments in education?*

The main conclusion from our empirical study supports the hypothesis that a positive relationship between a skewed sex ratio and males' years of education exists. However, unexpectedly, we found an even bigger positive correlation between an imbalanced sex ratio and females' educational attainment. This was not the main focus of this study and came as a somewhat unexpected result. Both results are based on the positive correlation in our analysis, but are not sufficient proof of a causal relationship.

Returning to our hypothesis, our empirical analysis shows that, based on our datasets, a skewed sex ratio has a positive effect on the overall educational attainment in China, for both males and females. Both the OLS and IV regressions reported a positive correlation and statistically significant results for the variable "Sex ratio for cohort 16-35" in both the composite sample and the samples split by gender, although column (1) in table (7) of the male sample reported a significance level above the conventional threshold. These findings indicate support for the hypothesis that males in areas facing more imbalanced sex ratios attain additional years of schooling, which may signal high intellectual capabilities and future earnings to potential partners. This result was consistent when we applied the trimmed sex ratio, which confirms that these findings were not driven by outliers. Further, incorporating province fixed effects in our OLS regressions had a minimal effect on the significance levels of our coefficients, but improved the explanatory value of the model (resulting in larger values of R-squared). The increase in R-squared was expected, as more variables were included. However, in order to detect the true additional explanatory value of including fixed effect controls, adjusted R-squared would have been a more appropriate measure.

Although, an increasingly skewed sex ratio has a positive effect on educational attainment, the economic significance of this finding is questionable, as the sex ratio variable consistently assumes small numerical values. In column (4) of the main OLS regression (table (4)), a one unit increase in the sex ratio would result in 0.06 additional years of schooling (corresponding to approximately 23 additional days)[19]. One potential explanation for this might be that it is not the skewed sex ratio per se that drives the differences in educational attainment, but rather that areas displaying higher levels of sex ratio imbalance might exhibit other similar characteristics as well. For instance, fierce labor market competition could result in individuals having to improve their relative labor market competitiveness by investing in additional education, regardless of gender. As discussed by Becker (1993), education facilitates access to higher paying jobs and reduces time spent as unemployed, which could potentially explain why areas with higher gender imbalance report more years of education. Additionally, China has experienced a tremendous economic growth in recent decades.

[19] 0.0626 * 365 = 22.85

This has led to an overall development of the Chinese society, with an increase in white collar jobs, which typically require higher formal education than traditional agricultural and manufacturing positions. Together with continuous educational reforms implemented by the state, this has resulted in overall general educational levels rising. The fact that areas with larger gender imbalances attain additional education could, thus, be a reflection of these areas "catching up" on other regions in terms of overall development and economic growth, resulting in a growing labor market and subsequently an increased demand for educated workers. In order to properly analyze this development, we encourage future research to investigate this topic by taking this time component into account.

In addition, it is plausible that demographic differences have an effect on an individual's educational attainment. When we split the sample by hukou status (see table (5)), we found an interesting result. The sample consisting of agricultural hukou registered individuals showed similar results as the main tables (table (4) and table (7)); an increase in the sex ratio correlates positively with years of education. In contrast, the non-agricultural hukou sample reported that an increase in the sex ratio skewness correlates negatively with years of education. One possible explanation for this might be that amongst individuals with non-agricultural registrations, the level of education is already considerably higher (see figures (5) & (6) in appendix). Since individuals are unlikely to marry outside their own hukou status, this means that additional educational attainment might not differentiate individuals to the same extent amongst non-agricultural individuals (Huang, 2001). As a result, these individuals may use other strategies to improve their standing in the marriage market. Following the reasoning from the previous paragraph, this pattern also could be driven by the rapid economic growth of the Chinese economy. As the agricultural registered hukou individuals in our sample obtain lower levels of education (see figure (5) & (6) in appendix) than non-agricultural individuals, their increased educational attainment might reflect a diminishing educational gap across the country, rather than a measure that agricultural individuals invest in education, to improve their marriage market competitive position when the sex ratio imbalance increases (which would suggest that the relationship between the sex ratio and educational attainment demonstrates a case of positive correlation, but not causality). In our main OLS table (4), as the sample consists of a majority of agricultural respondents (>80 percent), the positive result of the "Sex ratio for cohort 16-35" variable is likely mostly driven by factors influencing agricultural hukou individuals. We encourage future researchers to test whether the disparity between agricultural and non-agricultural individuals remains with a more evenly distributed sample in terms of hukou status.

As mentioned, the most surprising result from this investigation was that an increasing sex ratio skewness resulted in males attaining a lower increase in education compared to females. This finding is consistent and statistically significant across the OLS and IV model. For example, in table (7) of the IV model, a percentage unit increase in the "Sex ratio for cohort 16-35" results in 0.08 years of additional education for males (in column (2)), compared to 0.23 years of education for females (in column (4)). When comparing the difference in sex ratio effect between genders, the economic significance is somewhat larger compared to the general sex ratio effects, as the difference in educational attainment corresponds to approximately 55 days[20].

[20] 0.23-0.08 = 0.15. 0.15 * 365 = 54.75

The fact that males attain less increases in education than females when the sex ratio imbalance rises is not unambiguously in line with previous literature. For example, Lafortune (2013) found that an increase in the sex ratio would increase males' educational attainment compared to females in the US. Additionally, previous literature pointed out that the son preference in China has led to lower investments in girls' education in rural China, both in terms of school enrolment and completed years of schooling (Wang, 2005). However, Oliveira (2018) found that reduced family sizes, due to the fertility limits imposed under the OCP, led to increased educational investments in girls, closing the educational gap between the genders. Our findings suggest similar results. As areas with more imbalanced sex ratios have been proven to have been subject to stricter fertility limits, the fact that these areas also display larger investments in girls' education is aligned with the reasoning proposed by Oliveira (2018). As males in our sample on average inhibit more years of education, the skewed sex ratio results in a closing of the gender gap in educational attainment between the sexes here as well. Similar to the reasoning in previous paragraphs, one reason for this could be that areas with imbalanced sex ratios also display other similar characteristics. It is plausible that areas facing more skewed sex ratios and thus a stronger son preference, also disadvantage females in the labor market. Due to this inequality, females might have to attain additional education in order to compete with males for the same labor market opportunities which could explain why they attain more additional education compared to males when the sex ratio imbalance rises.

Further, Oreopoulos and Salvanes (2011) suggests that males' attractiveness on the marriage market increases with education. However, there might be other alternatives that make an individual an even more attractive candidate. As individuals' financial status is an important determinant of their marriage competitiveness, males facing fierce marriage market competition might consider the opportunity cost of attaining additional education (Becker, 1993). This pattern might be spurred by the relatively young average marriage age of 24 (as shown in table (1)), an individual would just have finished their master's degree at 24 and would not have been able to acquire significant wealth) which could motivate males to substitute potential higher life-time earnings for more short-term earnings opportunities, thus improving their marriage market outlooks. Further, the economic development in China might also have expanded the availability of high paying jobs that do not require high levels of education, making these attractive alternatives for males seeking possibilities to acquire wealth at a young age.

Although our findings confirm the hypothesis that a skewed sex ratio increases males educational attainment, we found the surprising result that males in fact attain less additional education compared to females with increasing sex ratios. As the literature presents ambiguous findings on the sex ratio effect on educational investment differences among the genders, we are not able to fully motivate our empirical findings with insights from our theoretical models or literature review. The discussion is therefore based on speculations of potential mechanisms, which should be further explored in future studies.

A short note should be made on the control variables. All control variables in the OLS and IV model report positive coefficients, meaning that they correlate positively with years of education.

These results are in line with previous research. For example, other research has found that increases in the mother's level of education increases the child's educational attainment (Harding et al., 2015). Similarly, higher family income has shown to correlate positively with educational attainment as well (Blanden and Gregg, 2004). The fact that our control variables correlate with the dependent variable in the expected direction confirms that the structure of our model is appropriate for assessing determinants of an individual's educational attainment.

9.1.1 Further discussion of IV results

One result from the first-stage regression in table (6) worth commenting on is the coefficients on the financial penalties for violating the OCP. The variables report similar coefficients as previous research conducted by Wei and Zhang (2011a). These findings confirms that stricter financial penalties are associated with more sex selective behavior and hence, a more imbalanced sex ratio. Further, as our tests for validating the instruments indicate that we can reject the null hypothesis of weak instruments and cannot reject the hypothesis of exogeneity, this provides confidence that our specified IV model is suitable for instrumenting the sex ratio.

9.2 Limitations of our study

9.2.1 General limitations

Our dataset is subject to several limitations which are worth commenting on. The most significant limitation to our study is the small size of the sample that we are examining. The main sample consists of 8,579 individuals which corresponds to approximately 0.0006 percent of the Chinese population. However, similar datasets and sample sizes have been used in previous research. i.e (Zhang et al., 2021; Zhang and Wei, 2011a). The main issue with a small sample size is the challenge of representability; how well does the sample represent the overall population and what conclusions could reasonably be drawn from the limited data? From a more technical perspective, a small sample size results in a decreased precision of the estimated coefficients, since the variance of these increases. Hence, this study may not have enough statistical power to detect significant results in our sample that reflects the effect of gender imbalance on education in the Chinese population as a whole.

Another limitation is the extensive prevalence of "missing values" in the CFPS dataset. This raises questions on the overall quality of the CFPS methodology as the data is collected through personal interviews where "missing values" should have been limited. This issue does not only reduce the sample size, but also makes several useful variables unreliable, and thus problematic to include in our model. For example, ideally, we would have liked to include a control for an individual's intellectual ability, as it intuitively should correlate positively with an individuals' educational attainment. However, approximately 60% percent of these values were reported missing, and would have reduced our sample size significantly (see table (8) in appendix).

An additional limitation is related to the level of clustering. The CFPS dataset reports information on each respondent's province and county of residence and birth, where the total number of provinces in the sample corresponds to 30 and the number of counties to 162. However, CFPS reported counties under pseudo-codes, where the real county names were not available to us.

Consequently, it was only possible to match the different datasets in our study by province, although it would have been better to do it on a county level, as it could have increased the sample variation. In order to improve the variation in our sample, we instead created five-year age cohorts on a provincial level which resulted in 519 clusters (493 cluster for the trimmed sex ratio results).

9.2.2 Model limitations

One main issue lies in the structure of our model. When investigating educational attainment, one main determinant is the individual's age. Since our sample consists of individuals aged 16-35, and school is mandatory until the age of approximately 15, our model does not provide the "full life" picture of an individual's education. For example, a 16-year-old in our sample could only have pursued one extra year of education, while a 35-year-old could have pursued 20 additional years. However, it is essential to point out that by excluding a control for age, we remove an important determinant of an individual's educational attainment and risk having biased estimates (see appendix for regressions controlling for age).

However, we do not include age in our model for several reasons. Firstly, many provinces in China increased their fine rates and implemented a premium over time, resulting in higher fine rates and higher probability of a premium for younger age cohorts, thereby making age negatively correlated with the instruments. By including age in our model, the fine rate and premium reported inconsistent coefficients (both positive and negative although mostly negative) in the first-stage regression, which made them intuitively difficult to interpret. Furthermore, a negative correlation between the instruments and sex selective behavior would suggest that stricter fertility limits in fact reduced the gender imbalance, something that has not been found by previous research. Previous papers using the instruments provided by Ebenstein (2010) have not included a control for age which might be due to this reason. Secondly, including age in our model causes the Kleibergen-Paap rk Wald F-statistic in the first stage regression to report values below 10, indicating that our instruments are weak and, thereby, that they do not serve as good predictors for the sex ratio. Thirdly, in the second-stage regression, the p-value for Hansen J statistic report values below 0.1, meaning that we can reject the hypothesis that the instruments are exogenous at a 0.1 significance level (and at a 0.01 significance level when applying the trimmed sex ratio). Hence, due to the many problems associated with the inclusion of an age control, we chose to exclude it. However, it is essential to point out that by excluding a control for age, we remove an important determinant of an individual's educational attainment and risk having biased estimates (see appendix for regressions controlling for age).

Another limitation of this study relates to the statistical validity of our results. As discussed in section (7.1.2), there is a risk for endogeneity issues in our OLS estimates. Due to the limitations of our data and the scope of this study, it is fair to assume that we did not exhaustively control for all factors influencing males' educational attainment, resulting in biased estimates. We have addressed this problem by adopting an IV approach, trying to isolate the exogenous effect of the sex ratio. However, the IV approach is a complex statistical method and is not without pitfalls. There is still a risk of biased estimates if the instrument is correlated with omitted variables, and the bias can turn out much larger compared to an OLS model, if the instruments are weak. In order to test for weakness, we conducted the Kleibergen-Paap rk Wald F-statistic and could reject

the hypothesis of weak instruments. However, this could also be a matter of coincidence as both the weak instrument test and exogeneity test reported that the instruments were invalid when we included a control for age in our model. Due to this fact, one has to be cautious when assessing the validity of our IV approach.

Lastly, a possible limitation of our study is that our model fails to control whether the effect of increased education in areas with high sex ratio comes from the general trend in China of increased education, spurred by the economic growth and educational reforms. It might be the case that these areas are catching up with the rest of China and, therefore, reflect a larger effect.

9.3 Future research

As discussed in section (9.1), this study finds support for our proposed hypothesis that a skewed sex ratio influences males' educational attainment. However, one main flaw of this study is its limited sample size. Therefore, we encourage future research to test whether our findings hold when conducting the study on larger samples or comparing variations over time. This would provide more robustness to whether sex ratio imbalance actually increases educational attainment among males, and even more among females.

In addition, we encourage future research to build on and improve methods used in this study. For example, a similar dataset containing higher quality responses, could increase the precision of the estimates by allowing more variables to be controlled for in the model. Also, our method does not include a trend variable for the overall increased educational attainment in China. As education levels have increased steadily during the last years due to overall economic growth and educational reforms, it would be interesting to explore whether the effect of the sex ratio still prevails after including a control for this effect. Similarly, since our model fails to properly include a variable controlling for age, which intuitively is an important factor when assessing educational attainments, we encourage future scholars to examine this topic.

Lastly, since our theoretical frameworks and literature review cannot fully explain why girls attain more years of education in areas where the sex ratio imbalance increases, we look forward to future studies examining this topic further.

10. Conclusion

The aim of this study has been to empirically examine the effect of the increasingly imbalanced sex ratios on educational attainment for males in China. Based on previous literature, Becker's Human Capital Theory and Spencer's Job-Market Signaling Model, we hypothesized that males facing fiercer marriage market competition may attain additional education in order to increase their marriage market competitiveness.

We obtained our main analysis sample of individuals aged 16-35, by using data from the 2010 China Family Panel Studies together with data from the China Population Censuses from 1990 and 2000. In order to empirically test the sex ratio effect on males' education, we first estimated the linear relationship between the two by using the OLS regression methodology, where the

average sex ratio by province of birth was regressed on competed years of education in 2010. In addition, this sample was also further split by hukou status, as an individual's hukou provides access to the welfare system, including public schooling. Moreover, to address potential endogeneity problems associated with the OLS, we applied an IV approach where we split the sample by gender, using instruments proposed by Ebenstein (2010).

Our main results can be summarized as follows. We found a positive relationship between skewed sex ratios and increased educational attainment among men, regardless of statistical method (OLS or IV). This finding was in line with the proposed hypothesis which suggests that males may use educational attainment as a tool to signal attractiveness towards potential female partners. Although the correlation between the sex ratio and education was found to be positive and statistically significant, it is worth noting that, the economic significance is relatively limited, as the sex ratio variable consistently reported small numerical values. On the other hand, as the sex ratios varied significantly by province and cohort (ranging between 92.02 - 124.92 males per 100 females), these disparities could result in large differences in educational attainment between groups.

We did not specifically focus on the effect of a skewed sex ratios on females' educational attainment, but we found a somewhat surprising result: an increasing gender imbalance has a larger positive effect on females' increase in educational attainment than for males (although both are positive). When comparing the difference in sex ratio effect between genders in the IV model, the difference in educational attainment corresponds to approximately 55 days. Hence, the economic significance of these results are almost twice as big as those from the OLS, although still relatively small.

We acknowledge the limitations of the available data, primarily the relatively small data sample that were available to us and that the statistical methods used in our study might give rise to unreliable results through endogeneity issues. Moreover, we recognize the limitations to the dataset in terms of missing values which hindered us from including all appropriate control variables. However, the limitations of this thesis leaves room for improvement and opportunities for future research within this field. We highly encourage future researcher to further study the effect of skewed sex ratios and its implications for educational attainment as it can provide additional insights into factors affecting educational inequality between genders and regional differences in males' educational attainment across China.

References

ANDERSON, S., 2007. The Economics of Dowry and Brideprice. *Journal of Economic Perspectives,* 21(4), pp. 151-174.

ANDREWS, I., STOCK, J.H. and SUN, L., 2019. Weak Instruments in Instrumental Variables Regression: Theory and Practice. *Annual Review of Economics, 11*(1), pp. 727-753.

AYOROA, P., BAILEY, B,. CROSSEN, A. and GEO-JAJA, M.A., 2010. Education in China: The Urban/Rural Disparity Explained. *World Studies in Education,* pp. 89-113.

BAI, J., TANG, J. and XIE, Q., 2022. Does children's marriage matter for parents' mental health? Evidence from China**.** *Journal of Asian Economics, 78*.

BBC, 2021. China NPC: Three-child policy formally passed into law. *BBC News*.

BECKER, G., 1993. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.

BECKER, G., 1981. A Treatise on the Family: Enlarged Edition. *Harvard University Press.*

BLANDEN, J. and GREGG, P., 2004. Family income and educational attainment: a review of approaches and evidence for Britain. *Oxford Review of Economic Policy,* 20(2), pp. 245-263.

BRITANNICA, 2022. Education of China. *Encyclopedia Britannica.*

CAMERON, L., MENG, X. and ZHANG, D., 2019. China's Sex Ratio and Crime: Behavioral Change or Financial Necessity? *The Economic Journal,* 129(618), pp. 790-820.

CHANG, S., KAN, K. and ZHANG, X., 2020. *Too Many Men, Too Short Lives: The Effect of the Male-Biased Sex Ratio on Mortality.*

CHEN, X. and SILVERSTEIN, M., 2000. Intergenerational Social Support and the Psychological Well-Being of Older Parents in China. *Research on Aging,* 22(1), pp. 43-65.

CHEN, Y., HONGBIN, LI. and MENG, L., 2013. Prenatal Sex Selection and Missing Girls in China: Evidence from the Diffusion of Diagnostic Ultrasound. *Journal of Human Resources,* 48(1), pp. 36-70.

CHANG, L., WANG, Y., SHACKELFORD, T.K. and BUSS, D.M., 2011. Chinese mate preferences: Cultural evolution and continuity across a quarter of a century. *Personality and individual differences,* 50(5), pp. 678-683.

CHINA STATISTICAL YEARBOOK, 2021. Population and Its Composition. *National Bureau of Statistics of China.* Available: http://www.stats.gov.cn/tjsj/ndsj/2021/indexeh.htm

CFPS, 2022. China Family Panel Studies. *Institute of Social Science Survey.* Available: https://www.isss.pku.edu.cn/cfps/en/

COALE, A.J. and BANISTER, J., 1996. Five decades of missing females in China. *Proceedings - American Philosophical Society,* 140(4), pp. 421-450.

CUI, Y., LIU, H., ZHAO, L., 2019. Mother's education and child development: Evidence from the compulsory school reform in China. *Journal of Comparative Economics,* 47(3), pp. 669-692.

DI LEO, G. and SARDANELLI, F., 2020. Statistical significance: *p* value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European Radiology Experimental* 4(18)

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

EDLUND, L., LI, H., YI, J. and ZHANG, J., 2013. Sex Ratios and Crime: Evidence from China. *The Review of Economics and Statistics, 95*(5), pp. 1520-1534.

FOREIGN LANGUAGE PRESS, 1982. The marriage law of the people's republic of China. *The Fifth National People's Congress.*

FU, G., LI, Y. and FEI, X., 2018. The evolution of urban wedding consumption in China since 1970s. *Journal of Contemporary Marketing Science,* 1(1), pp. 163-175.

GIETEL-BASTEN, S., HAN, X. and CHENG, Y., 2019. Assessing the impact of the "one-child policy" in China: A synthetic control approach*,* 14(11).

GOODKIND, D., 2011. Child Underreporting, Fertility, and Sex Ratio Imbalance in China. *Duke university press,* 48(1), pp. 291-316.

GRUIJTERS, R.J., CHAN, T.W. and ERMISCH, J., 2019. Trends in educational mobility: How does China compare to Europe and the United States? *Chinese Journal of Sociology*, 5(2). pp. 214–240.

GUO, J., LIN, S. and GUO, Y., 2018. Sex Birth Order and Creativity in the Context of China s One Child Policy and Son Preference.pdf. *Creativity Research Journal,* 30(4), pp. 361-369.

GUPTA, D. and SHUZHUO, L., 1999. Gender bias in China, South Korea and India 1920-1990: Effects of war, famine and fertility decline. *Development and Change,* 30, pp. 619-652.

HARDING, J., MORRIS, P., HUGHES, D., 2015. The Relationship Between Maternal Education and Children's Academic Outcomes: A Theoretical Framework. *Journal of Marriage and Family,* 77(1).

HEBREW UNIVERSITY OF JERUSALEM. Fine Rates on the One Child Policy. *Department of Environmental Economics and Management.* Available: https://scholars.huji.ac.il/avrahamebenstein/links/fine-rates-one-child-policy

HESKETH, T., LU, L. and XING, Z.W., 2011. The consequences of son preference and sex-selective abortion in China and other Asian countries. *Canadian Medical Association journal = journal de l'Association medicale canadienne,* 183(12), pp. 1374-1377.

HU, M. and SUN, Z., 2021. Too educated to be married? An investigation into the relationship between education and marriage in China. *Journal of Family Studies.*

HUANG, Y., 2001. Gender, *Hukou*, and the Occupational Attainment of Female Migrants in China (1985–1990). *Sage Journals,* 33(2), pp. 257-259.

IPUMS, Data Cart, *Ipums international.* Available: https://international.ipums.org/international-action/variables/group?id=core

JAYACHANDRAN, S., 2015. The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics, 7*(1), pp. 63-88.

JIANG, Q., LI, S. and FELDMAN, M.W., 2011. Demographic Consequences of Gender Discrimination in China: Simulation Analysis of Policy Options. *Popul Res Policy Rev,* 30, pp. 619–638.

JIANG, Q. and ZHANG, C., 2021. Recent sex ratio at birth in China. *BMJ Global Health,* 6(5).

JIN, X., GUO, Q. and FELDMAN, M.W., 2015. Marriage Squeeze and Intergenerational Support in Contemporary Rural China: Evidence from X County of Anhui Province. *The International Journal of Aging and Human Development, 80*(2), pp. 115-139.

KAPOOR, A. and DEBROY, B., 2019. GDP is not a measure of human well-being. *Harvard business review.* Available: https://hbr.org/2019/10/gdp-is-not-a-measure-of-human-well-being

LAFORTUNE, J., 2013. Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market. *American Economic Journal: Applied Economics,* 5(2), pp. 151-178.

LEE, D.S., MCCRARY, J., MOREIRA, M.J. and PORTER, J., 2020. Valid t-ratio inference for IV. *National Bureau of Economic Research.*

LI, J. and LI, J., 2021. Bride price and household income: evidence from rural China. *Journal of the Asia Pacific Economy,* pp. 1-20.

LI, W., SONG, C., XU, S. and YI, J., 2020. High Sex Ratios and Household Portfolio Choice in China. *The Journal of human resources, 57*(2), pp. 465.

MINISTRY OF EDUCATION, 2009. Compulsory Education Law of the People's Republic of China. *Laws and Policies.* Available: http://en.moe.gov.cn/documents/laws_policies/201506/t20150626_191391.html

MORRISON, W.M., 2019. China's Economic Rise: History, Trends, Challenges, and Implications for the United States. *Congressional Research Service.*

OECD, 2016. Education in China - A snapshot. Available: https://www.oecd.org/china/Education-in-China-a-snapshot.pdf

OLIVEIRA, J., 2018. Birth order and the gender gap in educational attainment. *Review of Economics of the Household,* 17(3), pp. 775-803.

OREOPOULOS, P. and SALVANES, K.G., 2011. Priceless: The Nonpecuniary Benefits of Schooling. *The Journal of Economic Perspectives,* 25(1), pp. 159-184.

RITCHIE, H., and ROSER, M., 2019. Gender ratio. *Our world in data.* Available: https://ourworldindata.org/gender-ratio

SEN, A., 1990. More than 100 million women are missing. *The New York review.*

SPENCE, M., 1973. Job Market Signaling. *The Quarterly Journal of Economics,* 87(3), pp. 355-374.

STOCK, J. and YOGO, M., 2005. *Testing for Weak Instruments in Linear IV Regression.* New York: Cambridge University Press.

TAN, C.K.K. and XU, Z., 2020. Virtually boyfriends: the 'social factory' and affective labor of male virtual lovers in China. *Information, Communication & Society,* 23(11), pp. 1555-1569.

THE NOBEL PRIZE. Gary Becker Facts. Available: https://www.nobelprize.org/prizes/economic-sciences/1992/becker/biographical/

UNITED NATIONS, 2019. World population prospects 2019. *Department of Economic and Social Affairs.* Available: https://population.un.org/wpp/Download/Standard/Fertility/

WANG, C., 2012. History of the Chinese Family Planning program: 1970–2010. *Contraception,* 85(6), pp. 563-569.

WANG, W., 2005. Son preference and educational opportunities of children in China— "I wish you were a boy!". *Gender Issues,* 22, pp. 3-30.

WANG, X., 2021. Permits, Points, and Permanent Household Registration: Recalibrating Hukou Policy under "Top-Level Design". *Journal of Current Chinese Affairs* 2021, 49(3), pp. 269–290.

WEI, S. and ZHANG, X., 2011a. Sex Ratios, Entrepreneurship, and Economic Growth in the People's Republic of China.

WEI, S. and ZHANG, X., 2011b. The competitive saving motive: evidence from rising sex ratios and savings rates in China. *Journal of Political Economy, 119(3).*

WEI, S., ZHANG, X. and LIU, Y., 2017. Home ownership as status competition: Some theory and evidence. *Journal of Development Economics, 127*, pp. 169-186.

THE WORLD BANK. The Human Capital Project: Frequently Asked Questions. Available: https://www.worldbank.org/en/publication/human-capital/brief/the-human-capital-project-frequently-asked-questions#HCP2

WU, X. and TREIMAN, D.J., 2004. The Household Registration System and Social Stratification in China: 1955-1996. *Demography,* 41(2), pp. 363-384.

YEE, R.K.H., 2006. Impact of Reforms on the Labour System in China, Policy and Society. 25(1), pp. 157-177.

YI, Z., PING, T., BAOCHANG, G., YI, X., BOHUA, L. and YONGPIING, L. Causes and Implications of the Recent Increase in the Reported Sex Ratio at Birth in China. Population and Development Review. 19(2), pp. 283-302.

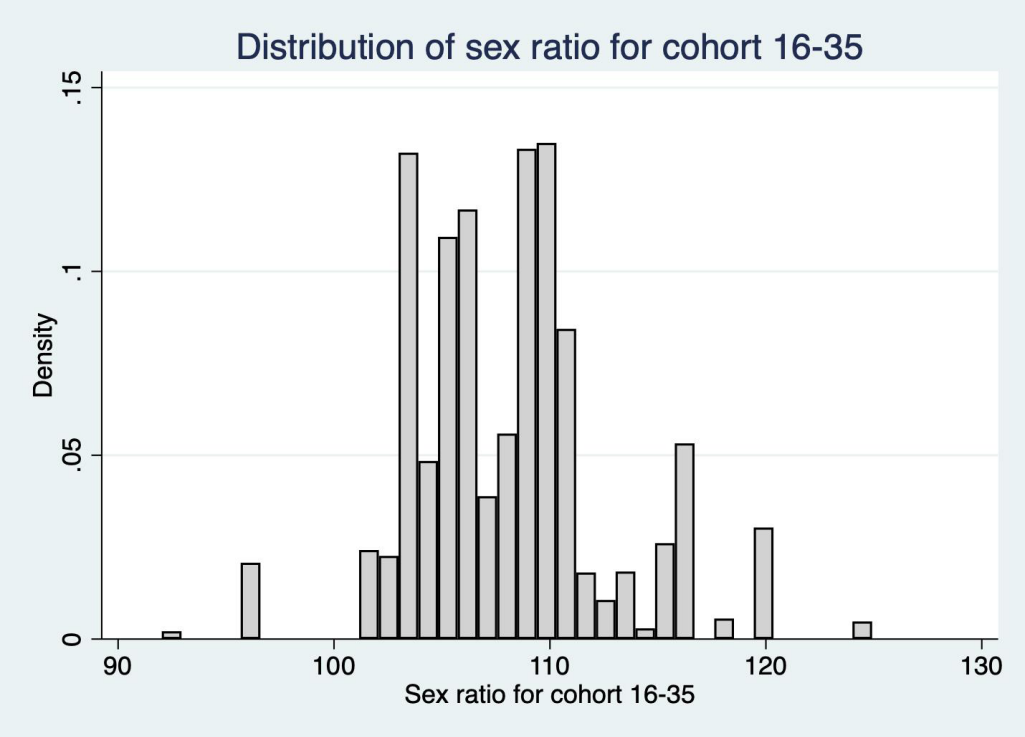
ZHANG, J., 2017. The Evolution of China's One-Child Policy and Its Effects on Family Outcomes. *Journal of Economic Perspectives,* 31(1), pp. 141-60.

ZHANG, K., HE, F. and MA, Y., 2021. Sex ratios and mental health: Evidence from China. *Economics & Human Biology,* 42.

ZHANG, L., 2022. Age matters for girls: School entry age and female graduate education. *Economics of Education Review,* 86.

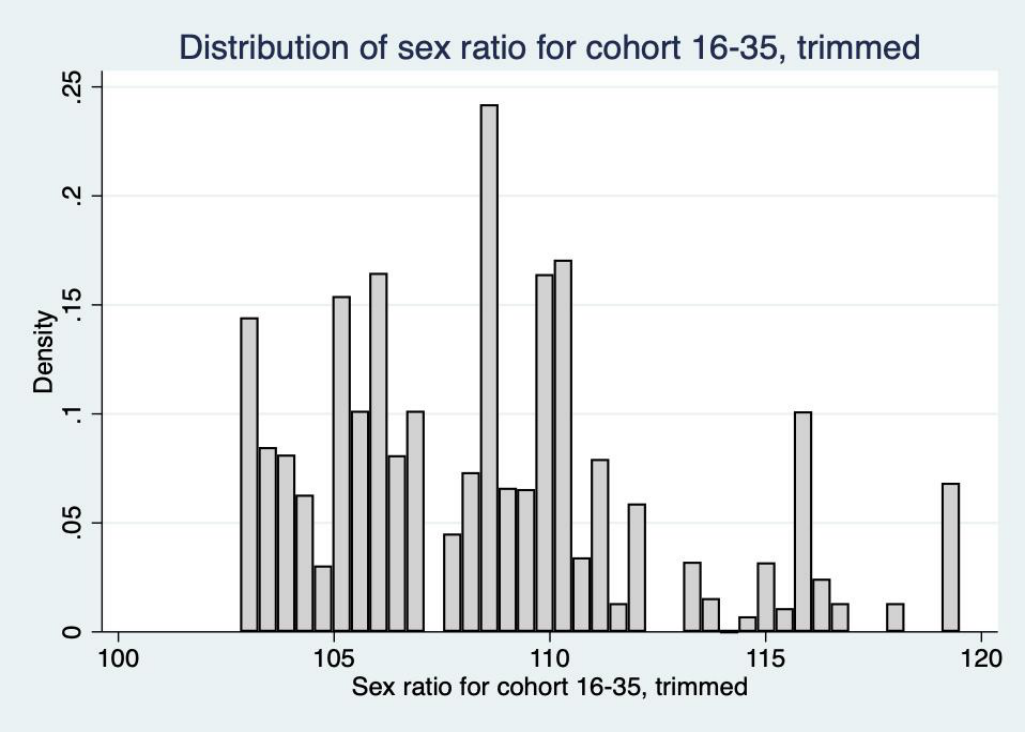
Appendix

Figure 3: Distribution of sex ratio for cohort 16-35



Source: CFPS (2022)

Figure 4: Distribution of sex ratio for cohort 16-35, trimmed values



Source: CFPS (2022)

Table 8 OLS Regression of Youth Sex Ratio on Years of Education in 2010, controlling for ability

	(1) Years of education	(2) Years of education	(3) Years of education	(4) Years of education	(5) Years of education
Male	2.622	2.703	2.175	2.151	3.959
	(2.029)	(2.078)	(2.010)	(1.982)	(2.557)
Sex ratio for cohort 16-35	0.00594	0.00744	-0.00440	-0.0160	-0.00335
	(0.0166)	(0.0165)	(0.0163)	(0.0174)	(0.0211)
Sex ratio for cohort 16-35 x male	-0.0251	-0.0240	-0.0201	-0.0196	-0.0362
	(0.0187)	(0.0191)	(0.0185)	(0.0182)	(0.0234)
Urban Residence	1.066***	1.015***	1.020***	1.109***	1.087***
	(0.119)	(0.120)	(0.114)	(0.124)	(0.127)
Hukou status at age 3	0.530***	0.515***	0.516***	0.566***	0.598***
	(0.0805)	(0.0813)	(0.0810)	(0.0871)	(0.0902)
Mother's level of education	0.589***	0.564***	0.524***	0.486***	0.463***
	(0.0625)	(0.0612)	(0.0608)	(0.0615)	(0.0626)
Log. of family income	0.611***	0.623***	0.605***	0.564***	0.603***
	(0.0723)	(0.0716)	(0.0705)	(0.0688)	(0.0711)
Log. of GDP per capita	0.473***	0.490***	0.451***	0.461***	0.454***
	(0.0609)	(0.0604)	(0.0595)	(0.0723)	(0.0801)
Math test	0.165***		0.139***	0.148***	0.151***
	(0.0169)		(0.0171)	(0.0160)	(0.0164)
Word test		0.351***	0.289***	0.269***	0.266***
		(0.0328)	(0.0348)	(0.0329)	(0.0338)
Constant	-5.468***	-6.428***	-5.382***	-3.515	-5.244*
	(2.089)	(2.052)	(2.050)	(2.377)	(2.901)
Observations	4,088	4,088	4,088	4,088	3,894
R-squared	0.333	0.330	0.353	0.378	0.381
Province FE	NO	NO	NO	YES	YES

Notes: Regression is identical to regression (4) but controlling for ability. Ability is measured as test score on a math- and word test. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS

Table 9 OLS Regression of Youth Sex Ratio on Years of Education in 2010, controlling for age

	(1) Years of education	(2) Years of education	(3) Years of education	(4) Years of education
Male	0.321***	5.373***	5.125***	4.971**
	(0.0902)	(1.927)	(1.884)	(2.287)
Sex ratio for cohort 16-35	-0.00435	0.0201	0.0142	-0.000180
	(0.0137)	(0.0161)	(0.0192)	(0.0222)
Sex ratio for cohort 16-35 x male		-0.0466***	-0.0442**	-0.0426**
		(0.0178)	(0.0173)	(0.0210)
Sq. Age	-0.000865***	-0.000860***	-0.000839***	-0.000975***
	(0.000263)	(0.000262)	(0.000238)	(0.000258)
Urban residence	1.230***	1.230***	1.318***	1.301***
	(0.109)	(0.109)	(0.113)	(0.116)
Hukou status at age 3	0.251**	0.248**	0.280**	0.279**
	(0.110)	(0.109)	(0.131)	(0.134)
Mother's level of education	0.925***	0.928***	0.856***	0.853***
	(0.0569)	(0.0568)	(0.0554)	(0.0568)
Log. of family income	0.675***	0.670***	0.657***	0.691***
	(0.0596)	(0.0596)	(0.0600)	(0.0626)
Log. of GDP per capita	0.690***	0.691***	0.535***	0.485***
	(0.0561)	(0.0562)	(0.0595)	(0.0631)
Constant	-6.147***	-8.775***	-6.198**	-4.303
	(1.850)	(2.039)	(2.624)	(2.975)
Observations	8,579	8,579	8,579	8,133
R-squared	0.317	0.318	0.350	0.354
Province FE	NO	NO	YES	YES

Notes: Regression is identical to regression (4) but controlling for age. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS

Table 10 First Stage Regression of Financial Penalties on Youth Sex Ratio, controlling for age

	Males		**Females**	
	(1)	**(2)**	**(3)**	**(4)**
	Sex ratio for cohort 16-35	**Sex ratio for cohort 16-35**	**Sex ratio for cohort 16-35**	**Sex ratio for cohort 16-35**
Fine rate	1.724**	-0.253	1.429**	-0.300
	(0.727)	(0.387)	(0.704)	(0.408)
Premium	-2.989***	-1.990***	-2.345***	-1.423***
	(0.495)	(0.443)	(0.467)	(0.435)
Sq. Age	-0.0120***	-0.0110***	-0.0115***	-0.0104***
	(0.000635)	(0.000536)	(0.000585)	(0.000528)
Urban residence	0.0724	0.0791	0.00376	0.0323
	(0.129)	(0.0951)	(0.109)	(0.100)
Hukou status at age 3	0.0780	0.0502	-0.0110	-0.0404
	(0.0719)	(0.0546)	(0.0294)	(0.0302)
Mother's level of education	-0.0603	-0.0776	-0.0674	-0.0131
	(0.0556)	(0.0502)	(0.0560)	(0.0421)
Log. of family income	-0.181***	-0.100**	-0.00745	-0.0266
	(0.0655)	(0.0471)	(0.0615)	(0.0457)
Log. of GDP per capita	0.180	-0.0580	0.0581	-0.0875
	(0.130)	(0.0645)	(0.0773)	(0.0654)
Constant	117.0***	120.7***	116.4***	119.4***
	(2.312)	(1.380)	(1.610)	(1.243)
Observations	4,155	3,924	4,424	4,209
R-squared	0.647	0.716	0.622	0.705
Kleibergen-Paap rk Wald F statistic	19.002	11.561	12.975	6.236
Province FE	YES	YES	YES	YES

Notes: Regression is identical to regression (6) but controlling for age. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

Source: CFPS (2022); IPUMS; Hebrew University of Jerusalem

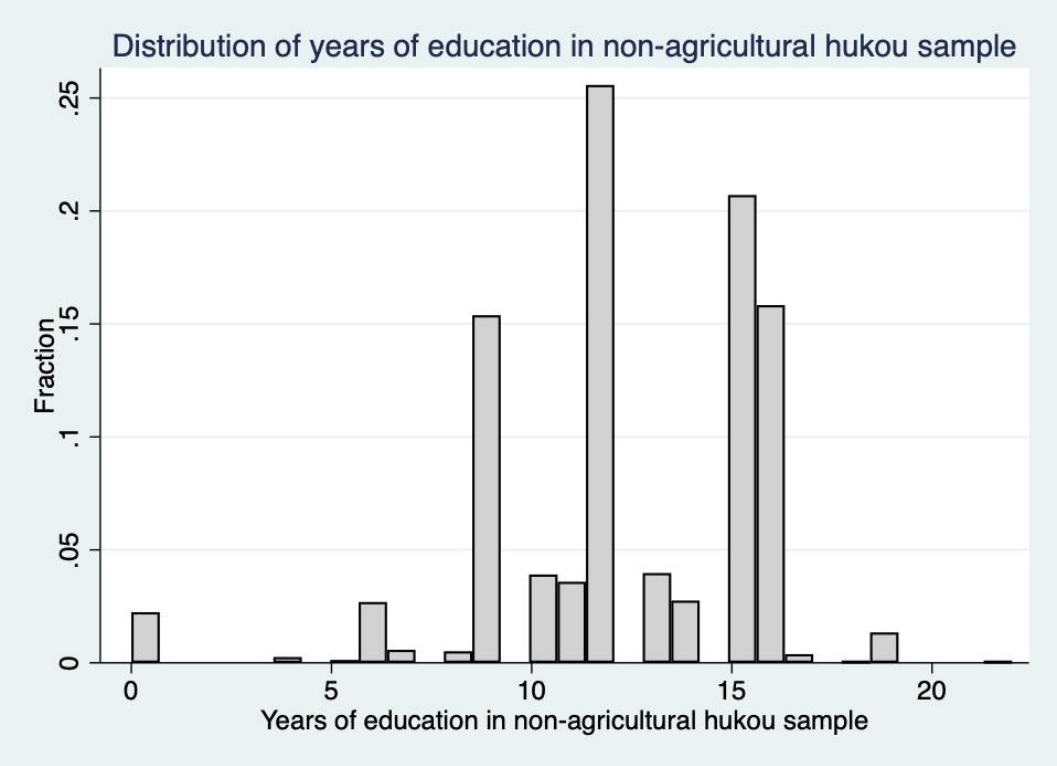
Table 11 Second Stage Regression of Youth Sex Ratio on Years of Education in 2010, controlling for age

	Males		**Females**	
	(1)	**(2)**	**(3)**	**(4)**
	Years of education	**Years of education**	**Years of education**	**Years of education**
Sex ratio for cohort 16-35	-0.172*	-0.370**	-0.0658	-0.301
	(0.0897)	(0.163)	(0.104)	(0.206)
Sq. Age	-0.00205**	-0.00367**	-0.00212*	-0.00425**
	(0.000973)	(0.00153)	(0.00113)	(0.00189)
Urban residence	0.983***	1.024***	1.682***	1.636***
	(0.158)	(0.166)	(0.152)	(0.156)
Hukou status at age 3	0.479***	0.482***	0.179	0.176
	(0.0926)	(0.0960)	(0.118)	(0.120)
Mother's level of education	0.679***	0.650***	0.981***	0.979***
	(0.0625)	(0.0682)	(0.0704)	(0.0716)
Log. of family income	0.576***	0.611***	0.656***	0.679***
	(0.0766)	(0.0796)	(0.0813)	(0.0854)
Log. of GDP per capita	0.508***	0.427***	0.581***	0.507***
	(0.0922)	(0.0953)	(0.0839)	(0.0972)
Constant	17.30	40.55**	3.089	31.03
	(10.64)	(19.42)	(12.26)	(24.21)
Observations	4,155	3,924	4,424	4,209
R-squared	0.293	0.272	0.407	0.390
Hansen J statistic (overidentification test)	0.2602	0.4090		0.3433
Province FE	YES	YES	YES	YES

Notes: Regression is identical to regression (7) but controlling for age. Stars indicate statistical significance *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the province and cohort level and reported in parentheses.

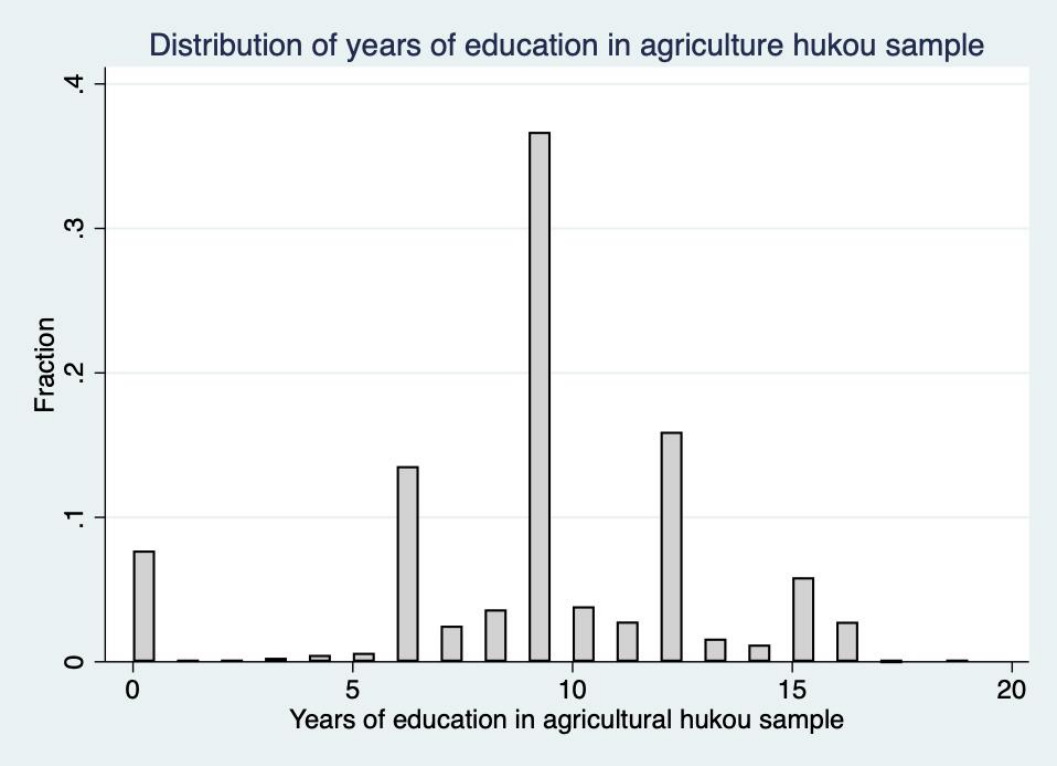
Source: CFPS (2022); IPUMS; Hebrew University of Jerusalem

Figure 5: Distribution of years of education for non-agricultural hukou sample



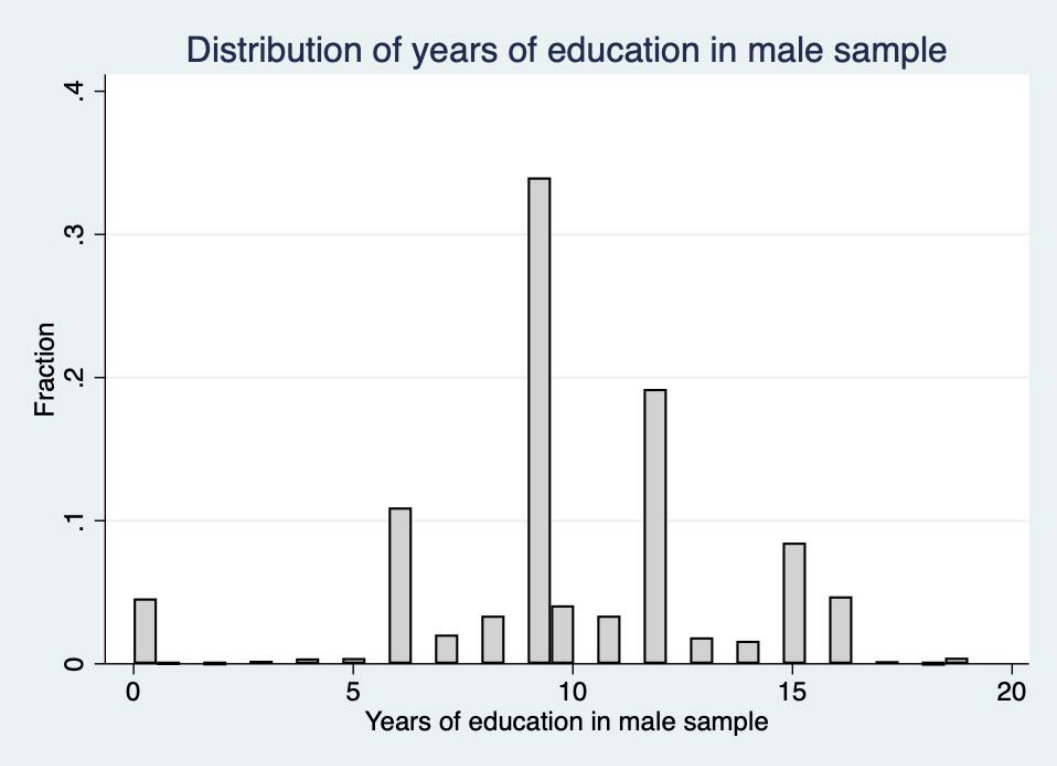
Source: CFPS (2022)

Figure 6: Distribution of years of education for agricultural hukou sample



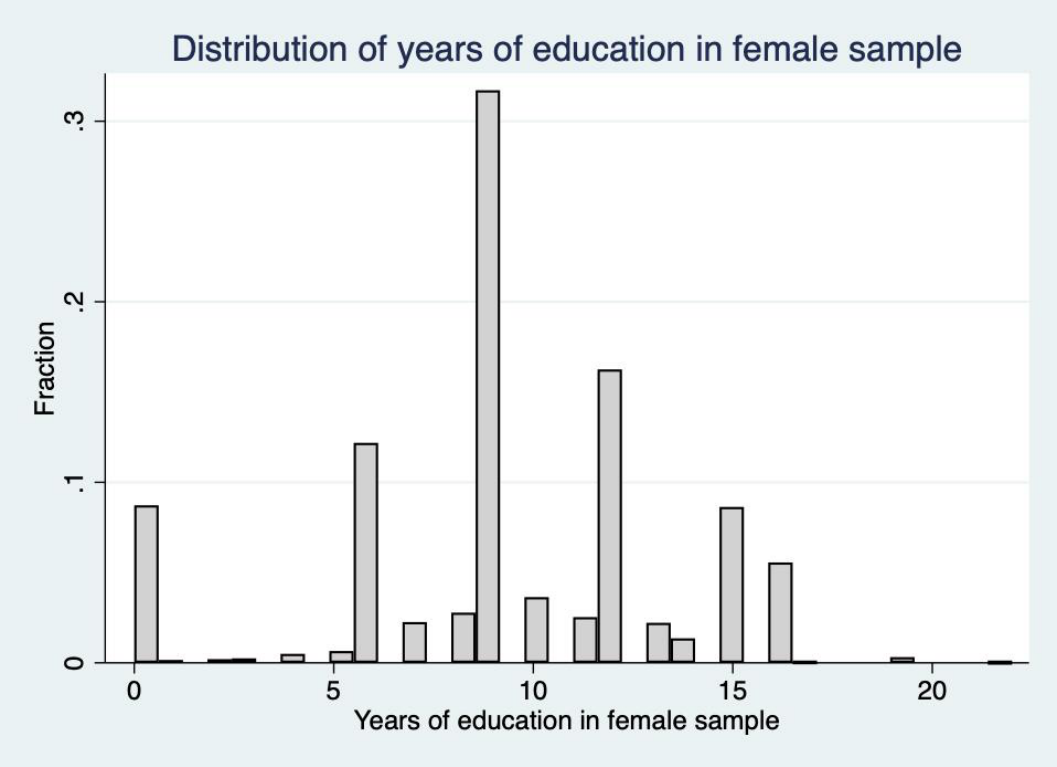
Source: CFPS (2022)

Figure 7: Distribution of years of education for male sample



Source: CFPS (2022)

Figure 8: Distribution of years of education for female sample



Source: CFPS (2022)

Table 12 Provinces of China

Beijing
Tianjin
Hebei
Shanxi
Inner Mongolia
Liaoning
Jilin
Heilongjiang
Shanghai
Jiangsu
Zhejiang
Anhui
Fujian
Jiangxi
Shandong
Henan
Hubei
Hunan
Guangdong
Guangxi Zhuang Autonomous Region
Hainan
Chongqing
Sichuan
Guizhou
Yunnan
Tibet
Shaanxi
Gansu
Qinghai
Ningxia Hui Autonomous
Xinjiang Uygur Autonomous Region
Total: 31

Source: CFPS (2022)