Essays on Childhood Environments and Human Capital Formation

Minhee Chae

A thesis submitted for the degree of
Doctor of Philosophy at
The Australian National University

December 2020

Chapter 2 is a joint work with Professor Xin Meng at the Australian National University and Assistant professor Sen Xue at Jinan University. My contribution to the chapter is 80 percent. Chapter 3 is a joint work with Professor Timothy J. Hatton at the University of Essex and Professor Xin Meng. My contribution to the chapter is a third. Except where otherwise indicated, this thesis is my own original work.

Minhee Chae
December 2, 2020

To my mother & pïanï

Acknowledgments

I sincerely thank Xin Meng for her supervision, encouragement, and endless support. I would also like to extend my appreciation to Alison Booth, Tue Gorgens, Tim Hatton, and Bob Gregory for countless lessons. I am also grateful for the support from the ANU Research School of Economics faculty, including, not limited to, Paul Chen, Dana Hanna, Fedor Ishakov, Timothy Kam, Juergen Meinecke, and Martine Mariotti. Many thanks to my Ph.D. fellows, friends, and housemates for the warmth and friendliness during the Ph.D. program. Finally, I would like to express sincere appreciation to my mother for her kind support.

Abstract

This thesis comprises three papers that examine childhood environment effects on individual cognitive and health outcomes in adulthood.

Chapter 2 examines the relationship between fertility and the gender gap in cognitive skills in a society where son preference is prevalent. Drawing on Becker's Quantity-Quality trade-off model, I empirically test if the trade-off between quantity and quality of children is larger for daughters than that for sons. To consider the endogenous nature of the demand for children, I exploit an exogenous variation in fertility due to China's family planning policy. We utilise the policy intensity information collected from hundreds of county gazetteers as an instrument for family size. The main results suggest that an additional sibling widens the gender gap in cognitive test scores by 33.2% of a standard deviation in the rural sample and 9.8% in the urban sample. The pattern is more pronounced in regions with a higher proportion of people who prefer a son over a daughter, and among households who face tighter budget constraints. I also provide suggestive evidence that our findings are strongly associated with belief in a son's role to carry on family lineage in Confucian tradition.

Chapter 3 explores childhood environment effects on adult height. Between 1950 to 1990, Chinese average adult height increased more than 1 cm a decade. This is an impressive achievement. This paper describes the trend of adult height growth during this period and explores the reasons behind such a trend. The result suggests that during the 40-year period, the growth of adult height increased, with the most substantial increase occurring in the last decade. One important reason behind the increase in adult height is the continued increase in government per capita spending on health and education throughout the forty-year period. However, the recent impressive growth was mainly due to the market-oriented economic reforms introduced in the 1980s. The observed height growth in the 1980s is a combination of two opposing factors, economic reforms and the introduction of the One Child Policy (OCP). We find that once we control for the 'number of siblings' after the introduction of the OCP, the revealed economic reform effects become much larger. The effect of OCP is, in general, negative on height and, in particular, for rural females owing to the widespread son-preference in the rural society.

Chapter 4 tests whether long-term exposure to single-sex schooling is beneficial for female cognitive performance. While recent studies have demonstrated that single-sex schooling improves girls' academic performance, little evidence is available as to how long it can be effective. By exploiting South Korea's random school assignment policy, we attempt to answer this question. The results suggest that attending single-sex schools improves female academic performance for the first three years of exposure, consistent with existing evidence. However, girls exposed to single-sex education for their entire secondary school life perform worse

than girls with less exposure at the college entrance exam. The estimates are robust to the inclusion of controls of individual, parental, and school characteristics. These patterns are not clear among boys. We provide suggestive evidence that show a gender-salient environment in single-sex schools may drive the results.

Contents

Acknowledgments **vii**

Abstract **ix**

1 Introduction **1**

2 Fertility and the Gender Gap in Cognitive Skills **3**

1 Introduction . . . 3
2 Background . . . 7
3 Data and Sample . . . 10
3.1 China Family Panel Studies . . . 10
3.2 Main sample . . . 11
4 Empirical Strategy . . . 16
5 Results . . . 22
5.1 OLS results . . . 22
5.2 IV results . . . 25
6 Channels . . . 28
6.1 Son preference . . . 28
6.2 Limited resources . . . 31
7 Robustness check . . . 33
7.1 Confounding factors . . . 33
7.2 Cognitive skill measure . . . 35
7.3 Birth order effect . . . 36
8 Conclusion . . . 38

3 Explaining Trends in Adult Height in China: 1950 to 1990 **39**

1 Introduction . . . 39
2 Background . . . 41
2.1 A centrally-planned economy 1949-1977 . . . 42
2.1.1 The establishment of urban healthcare system: 1949-1955 . . . 42
2.1.2 The planned economy and rural healthcare system: 1956-1958 . 42
2.1.3 The Great Leap Forward and the Great Famine 1958-1961 . . . 43
2.1.4 Recovery from the Famine 1962-1965 . . . 43
2.1.5 Cultural Revolution 1966-1976 and family planning policies 1973-1978 . . . 44

2.2 Market-oriented economic reforms and the One Child Policy 45
2.2.1 Economic reform and health care system evolution 45
2.2.2 One Child Policy and son-preference 47
3 Model specifications and estimation issues 48
4 Data and descriptive statistics 50
4.1 Data 50
4.2 Descriptive Statistics 52
5 Results 53
5.1 Long-term trends in the heights of Chinese 53
5.2 Height and its contributing factors 57
5.2.1 Socioeconomic factors 57
5.2.2 Economic reform, One Child Policy and height 60
5.3 The fertility effect 62
6 Robustness check 67
6.1 Self-reported height 67
6.2 Endogeneity of fertility 67
6.3 Missing values in local gazetteer data 69
6.4 Other potential channels 71
6.5 Other sensitivity tests 73
7 Conclusion 74

4 The Effects of Long-term Single-sex Schooling on Female Academic Achievement 75
1 Introduction 75
2 Background 77
2.1 Education in South Korea 77
2.2 The Equalisation Policy (EP) in Korea 78
3 Data 79
4 Empirical strategy 84
4.1 Endogeneity issue 85
4.2 Balance check 87
5 Results 90
5.1 Unconditional mean comparisons 90
5.2 Main results 92
6 Mechanism 95
6.1 Highlighting gender 95
6.2 Peer interactions 98
7 Robustness check 99
8 Conclusion and future research 102

5 Conclusion 103

I Appendix to Chapter 2 **105**

II Appendix to Chapter 3 **147**

III Appendix to Chapter 4 **165**

List of Figures

2-1 Fertility rate in China between 1950 and 1990 . . . 4
2-2 Gender differences in education and test score . . . 8
2-3 Number of siblings . . . 9
2-4 Gender differences in the standardised cognitive test scores . . . 15
2-5 The first instrument - Birth Control Rate (BCR) (%) . . . 19
2-6 The second instrument - Policy Fertility Index . . . 21
2-7 Son preference and sibling effects on gender gap in cognitive skills . . . 29
2-8 Son preference and sibling effects on gender gap in cognitive skills . . . 31

3-1 Grain consumption per person: 1950-1990 . . . 46
3-2 Height by rural-urban and gender: 1950-1990 . . . 47

4-1 Middle school performance comparison - (Left) Girls (Right) Boys . . . 90
4-2 High school performance comparison - (Left) Girls (Right) Boys . . . 91
4-3 High school performance comparison - (Left) Girls (Right) Boys . . . 92
4-4 High school Female teacher ratio . . . 96

2-A1 Main outcome variable - standardised vocabulary and math test scores . . . 106
2-A2 Test score distribution by test year and education group . . . 107
2-C1 Birth Control Rate in a county's gazetteer . . . 109
2-C2 Birth Control Rate collected from local gazetteers . . . 110
2-C3 Birth Control Rate histogram by year (Unit: individual) . . . 111
2-D1 The proportion of cases with end year observation . . . 116
2-D2 Survey-measured BCR and the actual gazetteer BCR . . . 117
2-E1 Actual values vs imputed values of Birth Control Rate . . . 119
2-E2 Actual values vs Imputed values of Birth Control Rate . . . 119
2-E3 Birth Control Rate at one's birth . . . 120
2-E4 Average Birth Control Rate within 1/3/5/7 years . . . 121
2-F1 How to measure the OCP target fertility at the prefecture level . . . 122
2-F2 Geographical distribution of the policy fertility . . . 123
2-F3 Policy fertility distribution in our sample counties . . . 124
2-F4 Province birth control policy information in Scharping (2003) . . . 124
2-F5 Fine rate distribution from Ebenstein (2010) . . . 128
2-F6 Ethnic minority share of women aged between 15 and 45 . . . 129
2-F7 County-level birth rates . . . 132

2-G1 Number of brothers and sisters 133

3-B1 WHO Child Growth Standards: median height-for-age 157
3-B2 Increments between consecutive ages 158
3-B3 Cohort distribution of the 'exposure' variable 159

4-A1 MSAT performance between excluded and included students 167
4-A2 Two middle school group comparison - MSAT performance - (Left) Girls (Right) Boys 168
4-A3 Two high school group comparison - CSAT performance - (Left) Girls (Right) Boys 171
4-A4 Three group comparison - CSAT performance - (Left) Girls (Right) Boys 172
4-A5 Three group comparison - MSAT performance - (Left) Girls (Right) Boys 173
4-A6 The relationship between performance differences and high school female teacher ratio 174
4-A7 Geographic coverage of the main sample 177

List of Tables

2-1 Summary statistics . . . 14
2-2 OLS results - Vocabulary test . . . 24
2-3 First stage results . . . 26
2-4 IV results - Vocabulary test . . . 27
2-5 Mechanism analysis - Limited resources . . . 32
2-6 Robustness check . . . 34
2-7 First born only . . . 37

3-1 Descriptive statistics . . . 54
3-2 Time trend in height growth . . . 55
3-3 Effects of regional social and economic factors on height . . . 59
3-4 Reform and OCP effects — Rural sample . . . 65
3-5 Reform and OCP effects — Urban sample . . . 66
3-6 Time trends in the CHNS and CFPS . . . 68
3-7 Instrumental variable results for fertility effect . . . 70
3-8 Other potential channels . . . 72

4-1 Main sample construction . . . 80
4-2 Summary statistics . . . 82
4-3 Number of students by four group in the main sample . . . 83
4-4 Random assignment details by cities . . . 86
4-5 Balance test . . . 89
4-6 Main results . . . 93
4-7 Female teacher ratio . . . 97
4-8 Randomness check: Single-sex middle school entrance . . . 101

2-B1 Main working sample construction . . . 108
2-C1 Birth Control Rate data availability by county . . . 111
2-C2 Number of observations available in gazetteer data . . . 112
2-D1 How to aggregate cases by prefecture and year . . . 114
2-D2 Contraception case distribution by reason . . . 115
2-D3 Birth Control Rate estimation sample in the 1988 Fertility survey . . . 115
2-D4 Missing values in contraception end year information . . . 116
2-E1 Estimation results used for imputation . . . 118
2-F1 Data used to predict the Policy Fertility Index . . . 123

2-F2 Incentive and Sanction rule information in Scharping (2003) . . . 126
2-F3 Average of Incentive and Sanction policy indicators in 1990s . . . 127
2-F4 Estimating α . . . 131
2-G1 OLS results - Rural sample (continued) . . . 134
2-G2 Years of schooling as the outcome variable . . . 135
2-G3 Sibling numbers by gender and relative birth order . . . 136
2-G4 OLS results - Math test . . . 137
2-G5 IV results - Math test . . . 138
2-G6 Proportion of women who prefer a son to a daughter . . . 138
2-G7 IV results by province . . . 139
2-G8 IV results using different year average Birth Control Rate . . . 140
2-G9 Birth order group dummy variables added . . . 141
2-G10 IV results including years of schooling . . . 142
2-G11 Son preference mechanism analysis . . . 142
2-G12 Vocabulary test score robustness check . . . 143
2-G13 Gender gap in NCEE performance . . . 143
2-G14 IV results . . . 144
2-G15 First stage results . . . 144
2-H1 Main sample selection process . . . 145
2-H2 Test procedure details by province . . . 145

3-A1 Main sample generation . . . 147
3-A2 Main estimation excluding 1950s cohort — Rural sample . . . 148
3-A3 Estimated results with total output . . . 149
3-A4 Selected results with Interaction terms . . . 150
3-A5 IV first stage results . . . 151
3-A6 Main estimation excluding predicted observations (Excluding one missing variable at a time) . . . 152
3-A7 Main estimation excluding predicted observations (Excluding all missing variables at once) . . . 153
3-A8 Estimated results with parental height variable included . . . 154
3-A9 Main estimation with three decade dummies . . . 155
3-A10 Main estimation with brothers and sisters . . . 156
3-C1 Data source . . . 161
3-C2 Data source and the proportion of missing values . . . 162

4-A1 Middle school types in Korea . . . 165
4-A2 High school types in Korea . . . 165
4-A3 Number of schools per school district . . . 166
4-A4 Middle school Female teacher ratio . . . 167

4-A5 Peer interaction in class 169
4-A6 Public high school only 169
4-A7 Early admission application 170
4-A8 Including non-metropolitan areas 170
4-A9 MSAT performance in the first three years - Girls 175
4-A10 MSAT performance in the first three years - Boys 175
4-A11 Main results using the raw test scores - Girls 176
4-A12 Main results using the raw test scores - Boys 176

Introduction

The importance of childhood environment on adulthood outcomes is well recognised in economics. Nevertheless, it is not well understood yet owing to difficulties in analysing the causal effects. The first challenge is the data limitation. Estimating the effects requires information on childhood environment and adulthood outcomes both, which demands data spanning more than two decades. This is particularly difficult for studies answering the question for developing countries or periods in the past, where advanced technology or system are absent. The second challenge is the endogeneity problem. Due to the potential presence of omitted variable bias and reverse causality, one needs to exploit exogenous variations in the childhood environment variables, which is hard to capture in reality.

This thesis comprises three papers that examine three major childhood environment effects on individual cognitive and health outcomes in adulthood: household environment, local economic environment, and school environments. Each study addresses the above-mentioned two issues by exploiting exogenous shocks in the natural experiments and hand-collecting data from historical books and identifies the causal effects of childhood environments.

Chapter 2 explores how family size in childhood affects cognitive skills in adulthood. According to the Quantity-Quality trade-off model developed by Becker and Lewis [1973], family size and the average quality of children are negatively associated due to scarce resources. I approach it in the context of gender. The main hypothesis is that the negative effect of family size is larger for daughters than sons when parents prefer sons over daughters. In other words, an additional sibling may widen the gender gap in abilities due to female disadvantages in a society where son-preference is prevalent. A major challenge to estimate the family size effect is the presence of potential omitted variable bias as the demand for children is endogenous. This study exploits an exogenous shock on fertility introduced by China's family planning policy. I hand-collected data, from hundreds of county gazetteers, on each local government's fertility control intensity over time and use it as an instrument for family size. The main results suggest that an additional sibling widens the gender gap in cognitive test scores 33.2% of a standard deviation in the rural sample and 9.8% in the urban sample. The pattern is more pronounced in regions with a higher proportion of people who prefer a son over a daughter, and among households who face tighter budget constraints. We also provide suggestive evidence that our findings are strongly associated with beliefs in a son's role to carry on family lineage in Confucian traditions.

Chapter 3 studies how local economic environments in early childhood are associated with physical development in adulthood. Between 1950 to 1990, Chinese average adult height has increased more than 1 cm a decade. This is an impressive achievement. This chapter explores the association between adult height and local environment changes during the period and examines to what extent it explains such a trend. The challenge for the analysis is a lack of data available on local economic conditions between 1950 and 1990. From local gazetteers compiled by local historians, we have assembled a unique set of data on county-level economic conditions, including economic output and local government social expenditure information. We find that the growth rate of the adult height has been increasing over the 40 years, and it was the last decade where we observed the most substantial increase. One important reason behind the adult height increase is the continued increase in government per capita spending on health and education throughout the forty-year period. But the recent impressive growth was mainly due to the market-oriented economic reform introduced in the 1980s. The observed height growth in the 1980s is a combination of two opposing factors, the economic reform and the introduction of the One Child Policy (OCP). We find that once we control for the 'number of sibling' after the introduction of the OCP, the revealed economic reform effect become much larger. The effect of OCP is, in general, negative on height and, in particular, for rural females due to the wide-spread son-preference in the rural society.

Another important environmental factor is education. Chapter 4 discovers school environment effects on student academic performance. It examines whether long-term exposure to single-sex schooling is beneficial for female cognitive development. While recent studies have demonstrated that single-sex schooling improves girls' academic performance, little is known as to how long such exposure can be effective. However, estimating its effect is challenging because students in single-sex schools may have different unobservable characteristics such as high passion for study compared to students in mixed-sex schools. This study exploits South Korea's random school assignment policy, which is an exogenous variation in the probability of attending single-sex schools, and attempt to answer the question. The result suggest that single-sex schooling improves female academic performance, but for the first three years of exposure only. First, girls in single-sex middle schools perform better than girls in mixed-sex middle schools at the Korean, English, and Math tests during middle school. Second, girls in single-sex high schools also perform better than girls in mixed-sex high schools at the college entrance exam. Both results are consistent with existing evidence on the positive effects of single-sex education for female students. However, girls exposed to single-sex schooling both in middle and high schools for six years - during the entire secondary school life - perform as bad as girls in mixed-sex high schools. It suggests the potential adverse effect of long-term single-sex education among girls. The pattern is not clear among boys. The estimates are robust to the inclusion of controls of individual, parental, and school characteristics. We provide suggestive evidence that gender salience in single-sex environments may drive the results.

Finally, Chapter 5 summarises the key findings in each of the three independent chapters and discuss future research directions.

Fertility and the Gender Gap in Cognitive Skills

1 Introduction

Gender gaps in educational and labour market outcomes which favour men have been narrowed over the past decades in many developing countries. This study links this phenomenon to fertility decline, which happened over the period and documents that fertility is a crucial determinant of gender gaps in a society where son preference is prevalent.

The idea is rooted in the Quantity-Quality trade-off (Q-Q) model developed by Becker and Lewis [1973]. This model suggests that, with a limited household budget, parents invest less in each child as the number of children increases. Thus there exists a negative relationship between fertility and level of human capital investment for each child. This general theory assumes that parental investment is equally distributed among all their children. However, when parents have strong son preference, the effect of family size on human capital investment may vary between girls and boys. If parents prioritise boys in resource allocation owing to son preference, an additional sibling will disadvantage girls more than boys.

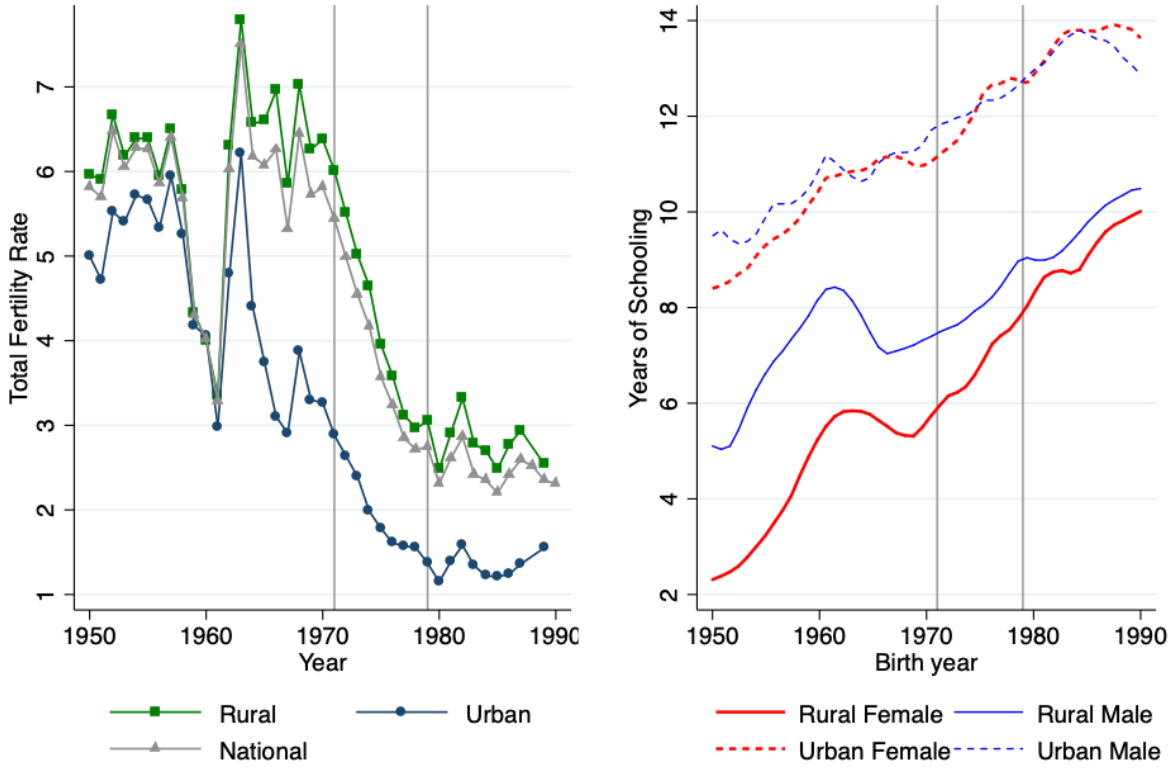
An additional sibling will reduce resources more for girls than for boys in the presence of son preference. When parents invest less on girls' education than on boys', an additional sibling will widen the gender gap in education and cognitive skills. The pattern should be more pronounced when son preference is stronger, in particular when they face tight budget constraints. With the same degree of son preference, parents who do not have enough resources to educate all their children well are more likely to disadvantage daughters to keep enough resources for boys than do parents who are well off.

We empirically test this hypothesis in the Chinese context. China is a society with a strong son preference [Jayachandran, 2015]. This preference is strongly influenced by the Confucian traditions that only sons can continue family lineage and perform ancestor worships [Milwertz, 1997]. The desire to have sons is also based on a traditional custom that sons provide old-age support for their parents [Das Gupta et al., 2003; Ebenstein and Leung, 2010]. A daughter is expected to be exported from her birth family to her husband's family upon marriage, which gives incentives for parents to invest less in a girl's education and more in a

boy's education. The cultural norm is described in the Chinese proverb, which goes "Raising a daughter is ploughing someone else's field" [Jayachandran, 2015].

China experienced a rapid decline in fertility over the past decades, as shown in Figure 2-1. At the same time, gender gaps in educational outcome and cognitive skills narrowed, which is in accordance with our main hypothesis.[1] The fall in fertility was mainly due to the Chinese government's family planning policy. It began in 1971 with the propaganda theme 'One Child isn't too few, two are just fine, and three are too many.', and the policy halved the total fertility rate between 1971 and 1976 [Scharping, 2003; Zhang, 2017]. For example, each local government kept detailed records on every women's menstrual cycles, past births, and contraceptive usage [Whyte et al., 2015]. In addition, the One Child Policy (OCP) further controlled fertility from 1979, although its impact on fertility remains questionable due to the market-oriented economic reforms that happened almost at the same time [Zhang, 2017; Garcia, 2019].

Figure 2-1: Fertility rate in China between 1950 and 1990



Our empirical strategy is to estimate the gender difference in family size effects on individual cognitive skills. We measure cognitive skills using individual word and math test scores provided by the China Family Panel Studies (CFPS), which is a longitudinal household and individual survey launched in 2010 by the Institute of Social Science Survey at Peking

[1] See Figure 2-2

University. The CFPS tests each adult respondent's cognitive skills and collects individual and household information, such as family size.

A major challenge to estimate family size effects on cognitive skills is the presence of potential omitted variable bias. Since family size and level of investment in children are both chosen by parents based on their preferences and circumstances, many of which are unobservable, the OLS estimation of family size on cognitive ability is inconsistent. Studies investigating the impact of family size on children's outcomes in China has often used the introduction of the One Child Policy (OCP) in 1979 as a natural experiment to identify the causal effect.[2] However, using the OCP as a natural experiment does not take into account the fact that before its introduction, China had already implemented an aggressive family planning policy, started in the early 1970s [Zhang, 2017]. Also, existing studies use the policy variations at either the national level or provincial level; however, the intensity of policy substantially varied by counties and prefectures [Scharping, 2003].

This study exploits the variation in the family planning policy intensity across different counties and over different years since serious family planning activities began in the early 1970s. We hand-collected data, from hundreds of county gazetteers, on each local government's records of contraceptive usage trends in the county over time and use this information as an instrument for family size. More specifically, our instrument is the proportion of fertile women who either were sterile or used contraception for each year and county. This variable is labelled as 'Birth Control Rate (BCR)' in local gazetteers, and ours is the first study to use this information as a measure of exogenous shock on fertility to our knowledge.

Our instrumental variable overcomes two major limitations in existing studies. First, it captures the intensity of the government fertility control policy since the early 1970s. Second, it varies at either county or prefectural level and also over years, while most of the existing studies rely on provincial-level variations.[3] We also construct the second instrument variable, which measures the number of children that each prefecture government targeted by year and use it as a robustness check.

Our main results suggest that an additional sibling widens the gender gap in the standardised vocabulary and math test scores by lowering female performance more than it does for male performance. The magnitudes are economically meaningful. One more sibling increases the gender gap in the vocabulary test scores by 33.2% of a standard deviation in the rural and 9.8% in the urban sample. This pattern is explained by sibling effects on gender gap in educational attainment to a great extent; however, we still observe the effect after controlling for years of schooling. It suggests that an additional sibling disadvantage affects girls in terms of not only the quantity of education but also the quality which is captured in test scores.

[2]Such studies include using provincial-level variation in the intensity of the implementation of the OCP as instrument [McElroy and Yang 2000; Liu 2014; Huang, Lei and Sun 2016]; using either the timing of the introduction of the OCP in a difference-in-differences setting [Li et al., 2011; Cameron et al., 2013; Qian, 2017] or in a regression discontinuity design [Qin et al., 2017]. Also, twin births are used to capture an exogenous shock on the family size under the OCP [Li et al., 2008; Rosenzweig and Zhang, 2009; Guo et al., 2018].

[3]Some studies use local variations in fine rates such as Liu [2014]; however, it only covers recent periods since 1990 due to data availability. On the other hand, our variable measures the policy intensity before 1990.

The mechanism analyses show that the female disadvantage of having siblings is more pronounced among individuals who grew up (1) in regions with a higher prevalence of traditional son preference and (2) in relatively poor households with tighter budget constraints. First, we measure the degree of son preference for each province using a question from the Women's Status in Contemporary China survey: *If only one child is allowed, would you like to have a son or a daughter?* Then we estimate the main model for each province and find that the sibling effects on the gender gap are larger among provinces with higher son preference. In addition, we show that a particular belief in which only sons can continue family lineage under the patrilineal system is the most associated with female disadvantage of having siblings rather than general gender role norms. Second, we run separate regressions for groups divided by parental education level, respectively. We find that the main results mainly come from individuals 1) who grew up in rural households and 2) whose parents are less educated than the average parents within one's birth cohort in the urban sample.

An important contribution of this study is the external validity of our results. First, to our best knowledge, this is the first study which utilises exogenous shocks on fertility since the early 1970s to test the Quantity-Quality trade-off theory. This policy shock decreased the national fertility from 6 to 2 children per woman before the OCP introduction, while most existing studies rely only on a marginal change from two children to one child enforced by the OCP. It enables us to generalise our findings to other developing countries with a high fertility level. Zhang [2017] also claims that a developing country may refer to the family planning policy implemented in the 1970s rather than the OCP. Second, the main sample is nationally representative. It includes individuals from 25 provinces in China, which covers most of the provinces.[4] Third, we measure cognitive skills in adulthood, in addition to the completed years of schooling. This measure may more directly relate to individual educational and labour market outcomes in adulthood. Additionally, this paper closely looks into the differential sibling effects between genders and provide suggestive evidence as to what cultural norms or beliefs could drive female disadvantage in a family.

The rest of the paper is organised as follows. Section 2 provides the background of the gender gap in China and fertility decline for the last decades. Section 3 introduces the data, and the main sample for our main analysis. Section 4 describes the empirical strategy with the instrumental variable construction. Section 5 presents the main results. Section 6 provides the potential channels of our main findings, and Section 7 show the robustness check results. Section 8 concludes.

[4]Many existing studies on the Quantity-Quality trade-off effects in China use data from 9-10 provinces [McElroy and Yang, 2000; Liu, 2014; Qian, 2017; Kubo and Chaudhuri, 2017] or twin data [Li et al., 2008; Rosenzweig and Zhang, 2009].

2 Background

China is regarded as a society with strong son preference [Jayachandran, 2015], which refers to the attitude that sons are more important and more valuable than daughters [Clark 2000]. This preference is deeply rooted in the Confucian traditions that only a son can perform important tasks such as ancestor worship and perpetuate the family line [Milwertz, 1997]. The desire to have sons is also based on a traditional custom that sons provide old-age support for their parents [Das Gupta et al., 2003; Ebenstein and Leung, 2010]. Empirical evidence also indicate the prevalence of son preference in China. Giving a birth to a son increases the mother's bargaining power in the household [Fan et al. 2018] and the male-to-female sex ratios are substantially higher than the natural sex ratio in China [Coale, 1991; Klasen, 1994; Klasen and Wink, 2003], especially under a strict birth control policy [Li et al., 2011].

The prevalence of son preference may disadvantage daughters in parental resource allocation. For example, mothers may spend more time with sons [Bo, 2018] and breastfeed sons for longer durations[Jayachandran and Kuziemko, 2005]. Sons may receive more childcare time and more vitamin supplementation than daughters [Barcellos et al., 2014]. Daughters are less likely to be vaccinated than boys [Borooah, 2004]. In short, sons get more resources than daughters when parents prefer sons over daughters.

As a consequence, one may expect a male-biased gender gap in the overall abilities in a society with strong son preference because parental investments are highly associated with the child's cognitive outcomes [Cunha et al., 2019; Francesconi and Heckman, 2016]. In fact, the adult female illiteracy rate is significantly higher than the male illiteracy rate in China according to China census 1982, 1990, and 2000 [World Bank, 2006].

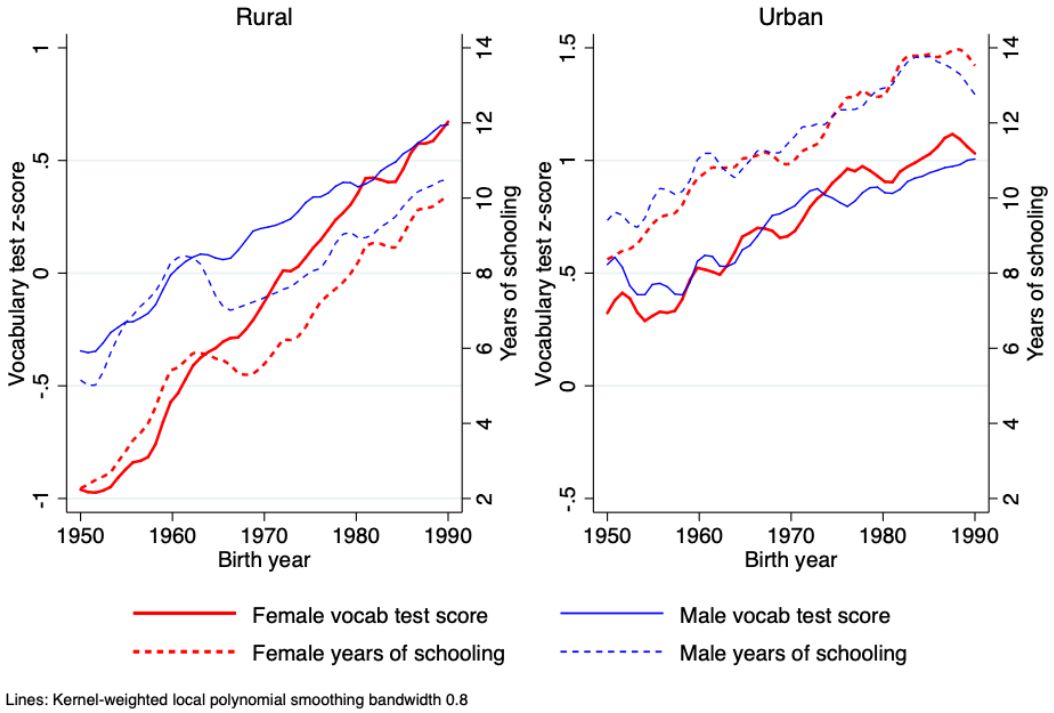
Figure 2-2 also shows men are more educated and perform better in standardised vocabulary test compared to women, especially among older generations and in rural regions where son preference is particularly strong.[5] Among 1950s birth cohorts, the gender gap in education years is approximately 3 years and the test score difference is 50% of a standard deviation. Men perform slightly better than women overall among urban people born before mid-1970. However, the difference is not significantly large among younger generations in terms of both test performance and schooling years. In fact, according to Lai [2010], girls outperformed boys in primary school and junior middle school in Beijing in the early 2000s.

Figure 2-3 indicates there had been a significant decline in the number of siblings over the past few decades in China, which is in accordance with the total fertility rate provided in Figure 2-1. Those born in 1950s have more than 3 siblings on average, while it decreases to about 1.5 in rural areas and less than 1 in urban China among those born after 1980.

These fertility trends are explained by the government policy history to a great extent. In the late 1950s and early 1960s, China experienced the Great Famine, which caused over 30 million deaths. Soon after, Mao launched the Cultural Revolution which threw the country into a political and economic upheaval. Agricultural productivity was stagnant over this

[5] See Section 3 for more details

Figure 2-2: Gender differences in education and test score



Source: China Family Panel Survey 2010 and 2014

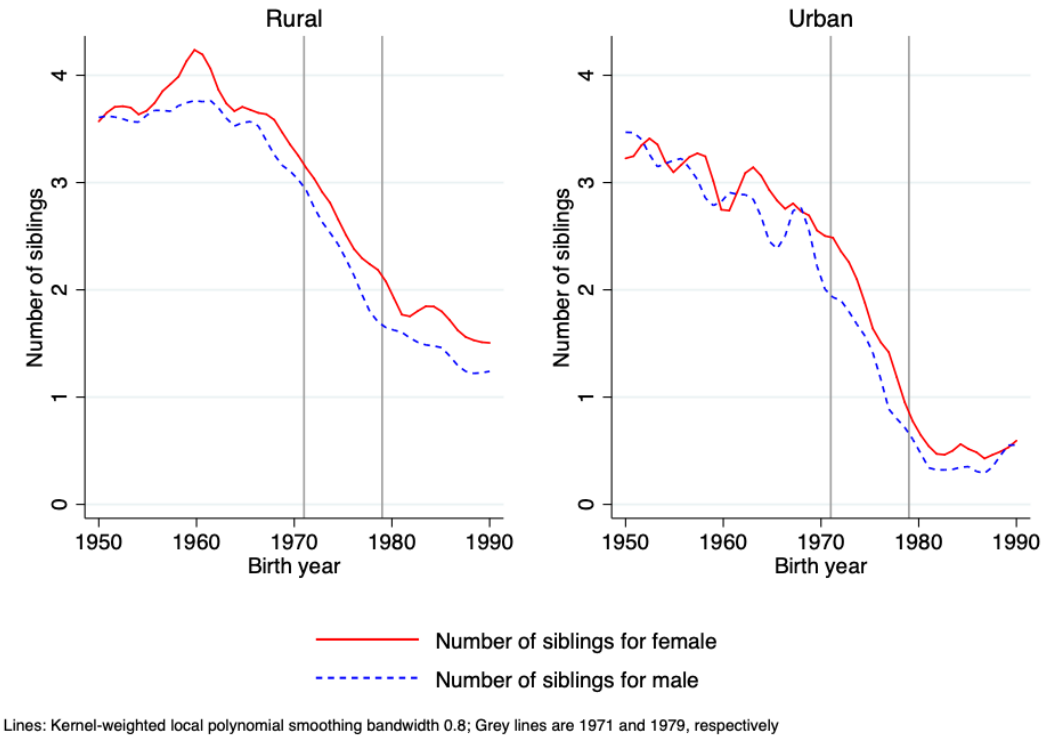
period. The clear memory of the Great Famine and the stagnation of the economy increased the fear that strong population growth might lead to a disastrous outcome. Moreover, high fertility which exceeded six births per mother throughout the 1960s concerned the government [Banister 1987].[6]

In the early 1970s, a debate over the possibility of a Malthusian Population Trap was initiated. Soon after, the Chinese government began a serious family planning campaign in 1971 with the propaganda 'One child isn't too few, two are just fine, and three are too many.' [Zhang, 2017]. Moreover, the Leading Group for Family Planning was officially established in 1973 and began the family planning policy, which encouraged people to have fewer children under a new slogan 'Later, Longer and Fewer' [Center for Population Studies, 1986; Peng, 1991; Feeney and Wang, 1993; McElroy and Yang, 2000; Cai, 2010; Ebenstein, 2010; Cameron and Meng, 2014; Whyte et al., 2015; Zhang, 2017]. 'Later' means late marriage and 'Longer' refers to longer birth spacing between children. More importantly, 'Fewer' means having no more than two children for urban families and three for rural families, and those who did not comply with the rule incurred penalties [Whyte et al., 2015].

This family planning campaign was technically voluntary. However, enforcement was coercive, which meant it did not simply rely on persuasion or voluntary compliance [Whyte et al., 2015; Zhang, 2017]. In fact, historical evidence suggests that many of the coercive en-

[6]The government attempted to control fertility in the period, but was not successful [Scharping, 2003].

Figure 2-3: Number of siblings



Source: China Family Panel Survey 2010 and 2014

forcement elements which are known as the OCP actually date from the 'Later, Longer and Fewer' campaign [Whyte et al., 2015]. Each local birth planning committee kept track of each fertile woman's menstruation, past births, and contraceptive usage. In rural regions, women who had three children were forced to get a sterilisation operation or insert an intrauterine device to stop further births. In some factories, female employees who became pregnant without permission were subjected to regular harassment to get an abortion. Also, above-quota births were denied household registration; in other words, they would be denied schooling and other government benefits [Whyte et al. 2015]. Statistical evidence also shows that the number of birth-control operations including female sterilisation and intrauterine device insertion in China began to sharply increase in 1971, and had more than doubled by 1975.[7]

As can be seen in Figure 2-1 and Figure 2-3, fertility fell significantly over this period. As a consequence, the population growth was cut by half between 1970 and 1976 [Scharping, 2003]. Total fertility rate decreased from more than 6 births per woman to around 2.8 births in rural areas and from 3 to less than 2 births per woman in urban areas between 1971 and 1978.

However, the nation was still facing food shortages, so the government decided to directly target the number of children per family. In December 1977, the State Planning Commission decided on demographic targets which further tightened previous birth control activities. In January 1978, the fertility control campaign entered a new era where a new policy of 'One

[7] See Figure 1 in Whyte et al. (2015) for more details.

is the Best and Two is the Most' and 'Reward Having One-Child and Punish Having Three' was introduced [Center for Population Studies, 1986; Peng, 1991]. This was then followed by a compulsory 'One-Child per Couple' policy.

As the main instrument of enforcement, economic and disciplinary sanctions were developed for unauthorised childbirth. In particular, for urban employees, wage deductions have always been the main sanction for violation of birth plans. Until the late 1980s, the majority of provinces deduct monthly wage by 10% as the punishment. There have been non-monetary sanctions as well. For example, those who exceeded the birth quota would not be promoted, the child of an out-of-quota birth would receive no medical insurance, and the family might not receive a housing allocation [Scharping, 2003]. In general, the policy has been more intensive in urban areas, and hence urban people experienced more intensive birth control. As an example, the penalties for unauthorised births are much more severe in urban areas than in rural [Bainster, 1987]. In some regions, urban people who violated the rule had to pay fines that were proportional to their monthly salaries as high as 70% [Li et al., 2011].

The OCP was strictly implemented in both rural and urban areas in the initial period. However, it was more difficult to implement in rural areas owing to strong resistance among many rural families, especially those with only one female child [Zhang, 2017]. Considering the circumstance, the central government finally relaxed the policy in 1984 and allowed families to have a second child if the first one was female [Scharping, 2003]. Later, a second child was generally allowed in many rural areas [Rosenzweig and Zhang, 2009].

However, it is controversial whether the OCP helped in reducing total fertility. In fact, the OCP introduction was not followed by a significant further decline in fertility until the end of the 1980s [Whyte et al., 2015]. Also, a slight decline in fertility after 1978 could be explained by the rapid economic development [Zhang, 2017; Garcia, 2019].

To sum up, it is clear that China had already enacted an aggressive family planning policy in the early 1970s, and its fertility rates had already dropped sharply before the enactment of the OCP [Center for Population Studies, 1986; Peng, 1991; Zhang, 2017]. The OCP implemented to reduce fertility further in both rural and urban China.

3 Data and Sample

3.1 China Family Panel Studies

Our main data come from the China Family Panel Studies (CFPS) as the data provide unique information on individual cognitive ability in adulthood. The CFPS was launched in 2010 by the Institute of Social Science Survey at Peking University. As of July 2019, the 2010 baseline study and 2012, 2014, and 2016 follow-up studies (waves) had been released. In this study, we only use the 2010 and 2014 waves, as our main dependent variable was only available in those two waves. The sample of the CFPS is drawn from 25 provinces in China, which covers 94.5%

of the Chinese total population.[8] It uses the probability-proportion-to-size sampling with administrative units at province, county/district, and village/neighbourhood/community levels and socioeconomic status as the stratification variables. Households are randomly selected from each sampled community; see Xie and Lu [2015] for a detailed discussion of the sampling design.

The CFPS conducted vocabulary and math tests in 2010 and 2014. All respondents aged over 10 who were present at the time of the survey were subject to the tests. The CFPS has two separate surveys: an Adult survey and a Child survey. Those who are below 16 are surveyed through the Child survey questionnaires and those aged 16 or above are defined as adults so participate in the Adult survey. Since our analysis focus on cognitive skills in adulthood, we only use those in the Adult survey. Among the Adult survey respondents, 99.9% of all of the adult survey respondents (33,563 out of 33,600) participated in 2010 and 85.1% of them (31,597 out of 37,147) participated in 2014.[9] We use these test score data from 2010 and 2014 as the main outcome variables in our analysis. Also, the CFPS provides comprehensive information on individual and household characteristics such as the number of siblings, education history, parental background, marriage, work, and other relevant information.[10]

3.2 Main sample

We separately constructed two subsamples from the CFPS survey - urban and rural samples - to take into account the fact that China has had a rural-urban divide policy and different socioeconomic environmental changes over the past decades. We define urban and rural based on household registration (*hukou*) status in a person's early childhood. The urban sample includes individuals whose household registration was urban (non-agricultural) and the rural sample includes those with rural (agricultural) household registration at age 3. Government policy, such as the family planning, applies to the household based on their *hukou* rather than the current residential place.[11]

The original sample comprises 28,193 rural and 5,168 urban observations in the Adults survey in 2010, and 22,221 rural and 3,220 urban were still in the 2014 sample. In addition, 2014 also includes a refreshment sample of 5,461 rural and 681 urban observations. Because our main outcome variable is the cognitive test scores, we exclude those who did not attend cognitive tests either in 2010 or in 2014. To reduce measurement error, for those involved in

[8] The excluded provinces/cities/autonomous regions are Xinjiang, Qinghai, Inner Mongolia, Ningxia, Hainan, Macao, Hong Kong, etc.

[9] In this survey, an individual aged 16 and over is defined as an adult.

[10] Another advantage of using the CFPS is that it provides one's birth county and household registration information at 3 years old. These information are useful to measure the birth control policy intensity in one's early childhood, which is the main instrument for sibling size in our analysis. Further details are available in Section 4. Empirical strategy.

[11] The household registration system (*hukou*) in China was initially developed to perform vital registration, limit rural-to-urban migration, and impose effective political and social controls [Potter, 1983]. At present, each person in China has an official record (i.e. the household registration), recording date of birth, place of birth, the present place of residence, and so on. The place at which one is registered becomes the person's official place of residence [Goldstein, 1987; Goldstein and Goldstein, 1990; Potter, 1983; Li and Cooney, 1993]

both years' tests, which account for 67.69% of the final sample, we take the mean value of their test scores in both years. To increase our sample size, we also include individuals who only participated in the test in one of the years. For these individuals, we take the test score from the year the information is available. As indicated earlier, in both years only a limited number of people did not participate in the test, the exclusion in this aspect only reduced our sample marginally (2.3%).[12] Further, we exclude those who were born before 1950 and those born after 1990. This exclusion reduced our sample by 24.4%. There are also missing values for other covariates. Excluding them, our final samples consist of 21,641 rural individuals and 4,058 urban individuals. Table 2-B1 provides detailed information on our sample exclusion rules and step-by-step exclusion which leads us to the final samples.

Figure 2-A1 presents the standardised test score distributions in our sample. The vocabulary test consists of 34 Chinese characters drawn from the language textbooks used in primary and secondary schools, and there are 24 math questions.[13] The same question sets are used in both 2010 and 2014. The individual final score is the number of correctly answered questions. Thus, the maximum vocabulary score is 34 and math score is 24. We compute the standardised score by test year for each subject. As indicated earlier, we use individual mean of the two standardised scores if one attended the test in 2010 and 2014. For those who attended the test only once, their scores are from the year the tests were taken.[14] Overall the vocabulary scores are more evenly dispersed than the math scores[15] For the standardised math test scores there are two spikes which account for 67.62% and 71.05% of the urban distributions in 2010 and 2014, respectively. These two spikes also account for 33.40% and 38.03% of the rural distributions in 2010 and 2014, respectively. It suggests that the math test scores do not have enough variations and they may not be suitable for our analysis. For this reason, we mainly use the vocabulary test score information.

Table 2-1 provides summary statistics for the main variables used for our main analysis. Men perform better than women in both rural and urban samples on average. Urban men perform slightly better than women do by 1% of a standard deviation in the vocabulary test and by 0.05% in the math test on average. The gap is significantly larger among rural people. Men perform better than women by 34% of a standard deviation on the vocabulary test and by 39% of a standard deviation on the math test in the rural sample. Similar patterns are found in education attainments. Rural men are more educated than rural women by 1.83 years and urban men have completed 0.24 years more education compared to urban women. The differences are statistically significant at the 5% level or below.

On the other hand, parental education attainments do not have such a distinct gender difference. The gender differences in parental mean schooling years are less than 0.05 years

[12] Table 2-B1 indicates the number of individuals whose *hukou* information available is 39,503 in the working sample and among them, 38,562 have the test score observations.

[13] The 24 math questions include addition, multiplication, logarithms, trigonometric functions, and so on.

[14] The main results are not sensitive to (1) whether we take the average or maximum of individual standardised test scores and (2) whether we use the raw or standardised test scores or not.

[15] Figure 2-A1 provides the raw score distribution in 2010.

and not statistically significant at the 10% level. Among rural people, women's parents are slightly more educated than men's parents, which is the opposite of the patterns of individual schooling years that we find above.

Women have more siblings on average than men do, and particularly they have more brothers than men do. Rural women have 0.18 brothers more than rural men, and urban women have 0.17 brothers more than urban men, and the differences are statistically significant at 1% significance level. On the other hand, gender differences in the number of sisters are close to zero in both rural and urban samples. Also, women are less likely to be a single child in both samples. These sibling patterns are to be expected in a society where there are fertility controls and a prevalence of son preference. Families with a female child are more likely to have additional births until they have a boy, whereas families with a male child are likely to stop giving birth.

Table 2-1: Summary statistics

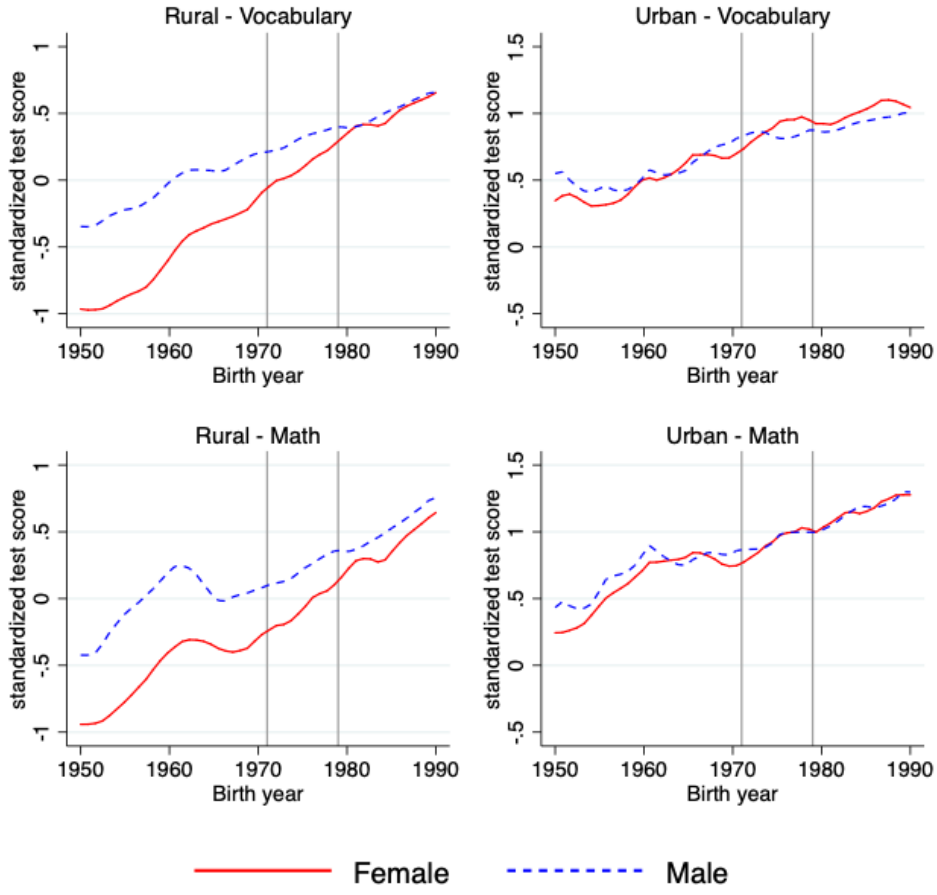
	Rural sample			Urban sample		
	Mean Female	Mean Male	Diff Female-Male	Mean Female	Mean Male	Diff Female-Male
Vocabulary test raw score (out of 34)	15.13	18.90	-3.78***	24.64	24.82	-0.18
	[10.60]	[8.89]		[6.89]	[6.31]	
standardised Vocabulary test score	-0.18	0.16	-0.34***	0.69	0.70	-0.02
	[0.97]	[0.81]		[0.63]	[0.58]	
Math test raw score (out of 24)	8.55	11.15	-2.59***	15.52	15.89	-0.37**
	[6.32]	[5.70]		[4.77]	[4.42]	
standardised Math test score	-0.22	0.16	-0.39***	0.81	0.87	-0.05**
	[0.94]	[0.85]		[0.71]	[0.66]	
Test Age	40.79	40.54	0.24	41.00	40.56	0.44
	[11.37]	[11.48]		[11.90]	[11.77]	
Schooling years[1)]	6.08	7.91	-1.83***	11.33	11.57	-0.24**
	[4.69]	[4.12]		[3.64]	[3.32]	
Parental schooling years	3.10	3.07	0.03	6.28	6.33	-0.05
	[3.25]	[3.20]		[4.19]	[4.29]	
Number of Siblings	3.04	2.79	0.26***	2.17	1.95	0.23***
	[1.85]	[1.86]		[1.90]	[1.80]	
Number of Sisters[2)]	1.51	1.51	-0.01	1.07	1.03	0.05
	[1.33]	[1.29]		[1.23]	[1.18]	
Number of Brothers[2)]	1.64	1.46	0.18***	1.16	0.99	0.17***
	[1.23]	[1.30]		[1.22]	[1.15]	
Number of Older siblings[2)]	1.62	1.63	0.01	1.21	1.18	-0.03
	[1.61]	[1.62]		[1.52]	[1.47]	
Number of Younger siblings[2)]	1.53	1.34	-0.19***	1.03	0.83	-0.19***
	[1.49]	[1.45]		[1.49]	[1.45]	
Number of Older sisters[2)]	0.79	0.86	0.07***	0.59	0.61	0.02
	[1.07]	[1.03]		[0.94]	[0.95]	
Number of Older brothers[2)]	0.83	0.78	-0.05***	0.62	0.57	-0.04
	[1.05]	[1.04]		[0.97]	[0.91]	
Number of Younger sisters[2)]	0.72	0.66	-0.06***	0.48	0.42	-0.06**
	[0.98]	[0.91]		[0.82]	[0.75]	
Number of Younger brothers[2)]	0.81	0.68	-0.13***	0.55	0.42	-0.13***
	[0.93]	[0.94]		[0.81]	[0.76]	
Birth order	2.54	2.46	0.08***	2.15	2.10	0.05
	[1.59]	[1.57]		[1.49]	[1.43]	
Dummy for one child	0.05	0.09	-0.03***	0.24	0.28	-0.04***
Dummy for current urban *hukou* registration	0.15	0.17	-0.02***	0.98	0.98	0.00
Dummy for currently living in urban regions	0.38	0.37	0.01**	0.95	0.95	0.01
Dummy for Han Chinese	0.91	0.91	0.00	0.96	0.96	0.01
Observations	10,974	10,667		2,032	2,026	
	21,641			4,058		

Note:
1) Years of schooling when one attended the test. For those who attended both in 2010 and 2014, mean years of schooling between 2010 and 2014 is used;
2) The sample size is 19,220 for rural and 3,719 for urban sample due to missing values;
3) Standard deviation in parentheses;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

Figure 2-3 summarises the number of siblings across birth cohorts in our sample. Sibship size started to decrease from around the late 1960s birth cohorts. In the urban sample, the number of siblings already almost reaches the lowest points before the legal introduction of the OCP in 1980 and remains at a similar level after 1980. However, for the rural sample, the sibling size continued to decrease after 1980 to a lesser degree. Overall, this supports the fact that the government's birth control policy has been intensive and effective since the early 1970s, and not only from the OCP implementation.

Figure 2-4 shows the gender gap in the standardised test scores in our sample across birth cohorts for rural and urban separately. The grey vertical lines indicates 1971 and 1979: the introduction of the first serious family planning campaign and the 'One-child per couple policies', respectively. Both test scores are standardised with zero mean and one standard deviation.

Figure 2-4: Gender differences in the standardised cognitive test scores



Standardized test scores with zero mean and one standard deviation;
Lines: Kernel-weighted local polynomial smoothing bandwidth 0.8
Grey lines are 1971 and 1979, respectively

The gender gap is narrowed among younger generations, especially in the urban sample. Women born after 1973 perform better than men in the vocabulary test and do as well as men in the math test. For the rural sample, however, women's progress relative to their male

counterparts has been continuous throughout the years. The early progress perhaps was made due to the introduction of gender equality after the communist party came to power in 1949. The Chinese communist government introduced its first marriage law in 1951 to stipulate that men and women are equal in marriage, family, land rights, education, and labour market participation [Niida, 1964; Yao and You, 2016; Booth et al., 2019]. Such gender equality policies had the most obvious effect in rural areas, which could have narrowed the gender gap. To consider this potential confounding factor, we will test the sensitivity of our results for the rural sample by excluding cohorts born before 1965.

4 Empirical Strategy

This paper tests whether fertility changes the gender gap in cognitive skills by changing the number of siblings. More specifically, we estimate gender differences in the sibling effects on individual cognitive skills in adulthood, which may be formulated as follows:

$$Y_{icts} = \beta_0 + \beta_1 Sib_{icts} + \beta_2 Sib_{icts} \cdot F_{ict} + \beta_3 F_{ict} + X_{icts} + \theta_c + \gamma_c l_t + v_{icts} \tag{1}$$

where Y_{icts} denotes individual standardised cognitive test scores at the survey year s for individual i who were born at birth county c in birth year t. Sib_{icts} is the number of siblings i has at the survey year s. We interact Sib_{icts} with a female dummy variable F_{ict} to capture the heterogenous sibling effects on individual test performance. X_{icts} represents a vector of individual characteristics, which includes parental schooling years, individuals' age, age squared, Han Chinese indicator, birth order and its interaction with female dummy variable. θ_c captures birth county fixed effects and l_t is a linear trends in Y for each county c. v_{ict} is the error term.[16]

Our main interest is the coefficient β_2. It captures the differential effect of the increase in the number of siblings on females relative to that of males. The negative coefficient indicates that an additional sibling generates a worse performance in the cognitive tests for females relative to that for males. Conversely, a reduction in one sibling improves female cognitive performance related to males.

However, the OLS estimations of β_1 and β_2 of equation (1) are inconsistent due to potential omitted variable bias problem. As we know, the number of siblings is the parental choice, which is related to parental preferences. In general, these preferences are unobserved and are often correlated with children's cognitive performance and educational outcomes. However, the direction of the bias, in this case, is unclear. It depends on the direction of the correlation between the omitted variable and the variable "number of siblings", and the children's outcomes. Qian [2017] discussed scenarios where the bias could be either positive or negative.

Existing studies exploit natural experiments to overcome this endogeneity issue. Angrist

[16] We additionally control for test year indicators - (1) attended only in 2010 and (2) attended only in 2014 - and birth county assumption indicators - (1) birth county information is used, (2) currently living county information is used.

and Evans [1998] use sibling gender composition as an instrument for the number of siblings. The idea behind this is that parents are more likely to have one more child if they have the same sex children only by assuming that parents prefer a mixed sibling-sex composition. Similarly, Lee [2008] uses the first child's sex as an instrument, as it may affect parental decision to have one more child, especially when they have son preference, and the first child is female. Twin births are also commonly-used instruments because this adds one more child that parents did not plan [Rosenzweig and Wolpin, 1980; Black et al., 2005; Rosenzweig and Zhang, 2009; Peter et al., 2018].

Many recent studies use China's OCP as a natural experiment to examine the causal effect of fertility on various outcome variables. Many of these studies rely on the timing of the introduction or revision of the OCP, either at the national level or provincial level as the instrumental variable [Qian, 2017; Huang et al., 2015]. Some use more detailed provincial policy information such as the magnitude of the fine imposed on people who violate the policy [Ebenstein, 2010; Huang et al., 2015]. However, this line of identification strategy may suffer from two potential problems.

First, as discussed before, the birth control policy was already significantly intensified since the early 1970s, and by 1979 (the year the OCP was introduced), the fertility rate dropped significantly [Zhang, 2017]. Thus, only using the introduction of the OCP as an IV may not allow the precise identification of the policy shock. Second, at the time of the introduction of the OCP, China also embarked on economic reform, which led to unprecedented economic growth. Therefore, the decline in fertility since 1980 may be attributed to economic development, not just the coercive enforcement of the OCP (see, for example, Whyte et al. [2015]). It is also noted that there had been substantial variations in the birth control policy intensity across counties or prefectures within each province [Scharping, 2003]. However, most of the existing studies rely on provincial policy information due to data limitation at the local level.

We construct two instrumental variables which resolve the above-mentioned limitations in the literature. Both instruments reflect the birth control intensity from the early 1970s. Also, the instruments vary across either counties or prefectures and birth years.

The first instrument we construct for the endogenous variable 'number of siblings' is the proportion of fertile women who use any contraceptive method at one's birth county for the period around the individual's birth. In the official government documents, this variable is labelled as 'Birth Control Rate (BCR)'. It is reported in most regional gazetteers and is used to reflect local government's fertility control outcome.[17]

We collect this BCR information from county/district and prefecture-level gazetteers ('local gazetteer' hereafter). According to Almond et al. [2019], local gazetteers are written by local historians and are not used for cadre evaluation, which reduces concerns over misreporting. Every gazetteer contains a chapter on 'birth control' and provides relevant information such as policy implementation history and its outcomes. Although each gazetteer follows their

[17] *jieyu* is translated as 'family planning' in the Chinese government documents. To clarify the direct meaning we call it 'birth control' in this paper.

format to provide information within the chapter, there are standard policy variables reported, including the BCR. Figure 2-C1 provides an example.

There is a total of 162 counties in our main sample, and we successfully collect the BCR information for 103 counties (63.6%) as plotted in Figure 2-C2, of which 60% are from county-level gazetteers and 30% from prefecture-level gazetteers as indicated in Table 2-C1. In the remainder of 59 counties, we are unable to find gazetteer information. We provide more details in Appendix C.[18]

We use the average rate over the period of three years before and three years after the individual's birth which is $\bar{B}_{ct}$ in equation (2).[19] $\bar{B}_{ct}$ captures the family planning policy intensity a few years around the timing of one's birth, which must be highly correlated with the number of siblings one has. More specifically:

$$\bar{B}_{ct} = \frac{1}{7}\sum_{t=-3}^{+3}\frac{CW_{ct}}{W_{ct}} \tag{2}$$

where $\bar{B}_{ct}$ is the average BCR within 7 years of one's birth year t in birth county c. CW_{ct} is the number of fertile women who either are sterile or use contraception and W_{ct} is the total number of fertile women in county c in year t. We average the proportion of CW_{ct} to W_{ct} between 3 years before and after one's birth including one's birth year. In other words, we use 7 year average of $\frac{CW_{ct}}{W_{ct}}$ for $t-3$, $t-2$, $t-1$, t, $t+1$, $t+2$, and $t+3$ as the instrument for the number of siblings of an individual born in t. The contraceptive method includes (1) female sterilisation, (2) male sterilisation, (3) intrauterine contraceptive device, (4) oral contraceptive pills, (5) abortions, (6) pills in external use, (7) injection, (8) condom, and (9) other medicine or medical machine. Intuitively, $\bar{B}_{ct}$ captures the likelihood of having siblings within 3 years before or after one was born.

To resolve this missing value problem, we use a regression imputation method. The method is to estimate missing values based on an imputation equation. It requires an assumption that the outcome variable's missingness is random conditional on observables included in the imputation equation. In our case, we estimate the following model:

$$B_{ct} = \beta_0 + \beta_1 SurveyB_{pt} + \theta_c + \gamma_t + \beta_2 X_{pt} + v_{ct} \tag{3}$$

$$t = 1972, 1973, ..., 1990$$

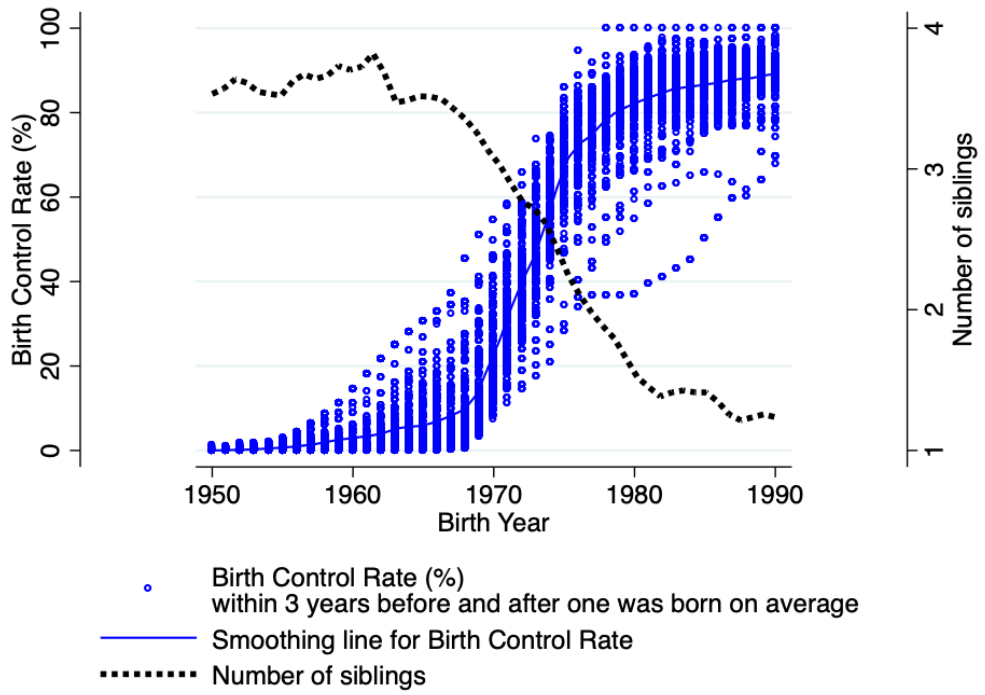
where B_{ct} is the actual BCR in year t for county c, $SurveyB_{pt}$ is the Survey-measured BCR, θ_c is county fixed effect, and γ_t is year fixed effect. X_{ct} includes a variable which indicates whether B_{ct} is prefecture level or not and its interaction with $SurveyB_{pt}$ in order to consider the unit difference between B_{ct} and $SurveyB_{pt}$. Intuitively, we predict the BCR missing values

[18] For Shanghai sample, we collect the BCR rates at municipal level for urban and rural areas separately from the Shanghai Population statistical yearbook. More details of data source and collection are available in Appendix C.

[19] We test the sensitivity of varying years of the coverage including 2 years before and after (=within 5 years of one's birth) and 1 years before and after (=within, and 2 year before and after).

using $SurveyB_{pt}$ using estimated coefficients. To generate $SurveyB_{pt}$, we use the Two-per-thousand fertility and contraceptive survey ('1988 Fertility survey' hereafter) conducted by the Chinese State Family Planning Commission in 1988. This survey collected information on each woman's fertility and contraception history among 459,269 ever-married women. The data consists of 544,190 contraception case observations including contraception type, start and end year, and the reason for each contraception. We (1) count the number of fertile women who use contraception by year and prefecture and (2) divide it by the total number of fertile women by year for each prefecture. Appendix D provides details on how to generate $SurveyB_{pt}$.[20] Then we regress the actual BCR on $SurveyB_{pt}$ and estimate the fitted values to fill missing values, as indicated in equation (3). According to the estimation results in Table 2-E1, our imputation model in equation (3) successfully explains 78% of variations in the actual BCR B_{ct}. Appendix E further provides evidence of the precision of imputation.

Figure 2-5: The first instrument - Birth Control Rate (BCR) (%)



Since the birth control policy was initiated in the early 1970s, we have only a few observations of the BCR from local gazetteers between 1950 and 1971. For this period, we use the Survey-measured BCR instead. Details of how to generate the Survey-measured BCR are available in Appendix D. Figure 2-5 plots the distribution of this first instrumental variable. In summary, first, we construct the BCR data using (1) the BCR from local gazetteers in 1972-1990 and (2) the Survey-measured BCR in 1950-1971. Figure 2-E3 shows how we combine those two variables. Second, using this data, we generate the average BCR variable within 3 years before and after one was born (= within 7 years including one's birth year; $\bar{B}_{ct}$) as in equation

[20] Due to limited number of observations available for each cell by county and year (10.76 observations per cell on average), we estimate $SurveyB_{pt}$ at prefecture level (104.74 observations per cell on average).

(2). Figure 2-E4 plots the average BCR within 7 years, 5 years, 3 years, and at one's birth.

The second instrument is an index that measures the number of children that each couple is allowed to have under the OCP at the prefecture level, which is called the Policy Fertility Index (PFI). That is, the PFI captures the number of children that each local government's birth control policy targets. We generate it using several data sources together with predictions for places and years where data are not available. Appendix F provides detailed information on how to construct this variable. We use the PFI as a sensitivity check of the main results, so the main analysis relies on the first instrument BCR.

The main data we use is the prefecture-level policy target births produced by Gu et al. [2007].[21] They collected information on fertility control policies introduced at 420 prefectures during the late 1990s. By merging this data with the 2000 Population Census, they estimate the number of children that each couple was allowed to have in the late 1990s, which summarises each prefecture's policy intensity (See Gu et al. [2007] for more details). This variable is called 'policy fertility'.

However, the policy fertility data are only available for the 1990s. Thus, we estimate the policy fertility for the years before the 1990s. To do so, we collected annual provincial-level data on the birth control policy details from Scharping [2003]. This data includes sanction and incentive rules for each province between 1979 and 1999. In total, there were 17 incentives and 25 sanction rules (see Figure 2-F4 and Table 2-F2 for details). We generate dummy variables for each of these incentives and sanction rules for each year over this period, where one indicates the policy was implemented and zero otherwise. We also collected information at the provincial level for different years the level of fines imposed on individuals if they violated the OCP published in Ebenstein [2010]. Using these provincial-level OCP policy data in the 1990s together with the prefecture-level policy fertility data for the same period from [Gu et al., 2007], we estimated the following equation:

$$P_{90s,f} = \alpha_0 + \sum_{k=1}^{K} \alpha_{1,k} S_{k,90s,p,u} + \sum_{l=1}^{L} \alpha_{2,l} I_{l,90s,p,u} + \alpha_3 F_{90s,p,u} + \alpha_4 M_{90s,f,u} + \epsilon_{90s,f,u} \tag{4}$$

where P is the policy fertility for prefecture f in the 1990s. S_k and I_l indicate the kth sanction rule and the lth incentive rule indicator at province p in the 1990s, and F is the fine rate for an unauthorised birth at province p in 1990s. We include 7 sanction and 3 incentive policy indicators for urban (i.e., $K = 7$ and $L = 3$) and 9 sanction and 8 incentive indicators for rural (i.e., $K = 9$ and $L = 8$).[22] $M_{1990s,f}$ is the share of the ethnic minority population (non-Han Chinese) of fertile women in the 1990s. We use this to capture the severity of the policy implementation as ethnic minorities are allowed to have more children in general [Cameron and Meng, 2014; Li et al., 2011]. This variable is generated using the 2000 Population Census

[21] The detail of this data source is presented in Appendix F.

[22] There are a total of 17 incentives and 25 sanctions. However, many are collinear with each other. We tested different specifications and discovered that the model is not sensitive to whether we include all of them or some of the most sensible ones. We tested what to include by estimating a single equation between the prefecture-level policy fertility and the incentive or sanction rule dummies one at a time. We kept those with sensible signs and coefficients.

data. The estimated coefficients are then used to predict for the years between 1979 and 1990 as we have data on all the explanatory variables between 1979 and 1990. For the years before 1979, we use county-level birth rate collected from county or prefecture gazetteers. See Appendix F for more details.

Figure 2-6: The second instrument - Policy Fertility Index

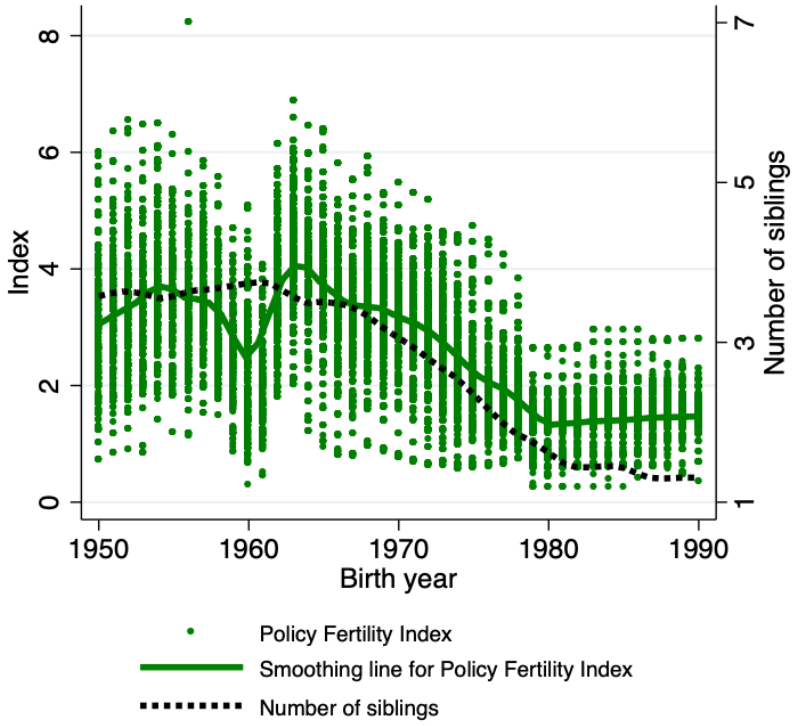


Figure 2-6 plots this PFI together with the number of siblings (on the left Y-axis) against birth year. It can be seen that the variable is highly correlated with the sibship size in our main sample.[23] However, this instrument has a few discontinuity points, while sibling size moves smoothly over time. This is because we used a specific year's fertility trend to generate the instrument, while sibling size reflects the fertility decision in multiple years. For this reason, we compute the seven-year moving average of this instrument in our analysis. More details of the construction process are available in Appendix F.

Valid instruments should be strongly correlated with the number of siblings and should not affect individual outcomes except through family size. One concern is that regional variations in the family planning policy enforcement may reflect family preferences and socioeconomic conditions of the region. If this is the case, our instruments could possibly be correlated with unobservable parental preference or household attributes. We control for county fixed effects to consider such differences in regional characteristics, which might affect the policy intensity. The proportion of women using contraception may capture individual decisions to control births rather than to comply with the government policy. Table 2-D2 shows that

[23]The correlation coefficient between sibling size and this instrument is 0.5152 for urban and 0.3747 for rural.

75.61% of 544,190 contraception cases reported in the Two-per-thousand fertility and contraceptive survey occurred due to the government policy.[24] The contraceptive usage proportion may not be highly correlated with family size if most of the contraception cases are temporary, such as condom usage. However, Table 2-D3 indicates 84.93% (=6.19+25.00+53.74) of 544,190 contraception cases are either sterilisation or intrauterine device, which requires a medical operation. We further discuss the relevance issue with the first stage results in the next section.

5 Results

5.1 OLS results

Panel A and B in Table 2-2 report the OLS results for rural and urban samples, respectively. Column 1 reports the results with sibling variable and a gender dummy variable only. It shows that an additional sibling is associated with a decrease in individual standardised vocabulary test scores by 12.4% and 9.8% of a standard deviation for the rural and urban samples, respectively. The coefficient on female indicator suggests that female test score in rural areas is 31.1% of a standard deviation lower than their male counterparts, but in urban areas, there is little gender difference in test scores.

Column 2 adds an interaction term between the number of siblings and the female dummy variable, which captures gender differences in sibling effects. After adding this interaction term, the negative coefficients of the number of siblings and the female dummy for the rural sample become much smaller (in absolute term) compared to those presented in column 1, whereas for the urban sample, the coefficient on the number of siblings hardly changed. These results suggest that in rural area, there exists a strong female disadvantage in having additional siblings. The gender gap in the test score among individuals without siblings is 14.8% of a standard deviation in the rural sample, and the gap increases by 5.6% of a standard deviation for every additional sibling.

Column 3 includes birth-county fixed effects. The coefficient on the number of siblings substantially reduced but is still statistically significant at the 1% level for both rural and urban samples. The gender difference in an additional sibling effect slightly increased for both rural and urban samples.

Column 4 adds parental education (to proxy for family economic background), Han ethnic indicator, and an indicator for the birth order and its interaction term with female dummy, age, age squared, and the county-specific time trend. While the coefficients on the female dummy variable and the interaction term between the number of sibling and female dummy remain negative and statistically significant, the coefficient on the number of siblings switches the sign to positive for the rural male. Thus, including the full set of controls, we observe that for males in the rural regions having an additional sibling increases their test score by 4.2%

[24] The data covers from 1950 to 1988.

of a standard deviation, whereas the effect for women is a negative 5.3% (=0.042-0.095) of a standard deviation. For the urban sample, the pattern on gender gap of having additional siblings is the same as that found in the rural sample, but the magnitude is smaller.

Interestingly, the gender gap for having an additional sibling increased substantially for the urban sample from column 3 to column 4. In column 3, the gender gap is not precisely estimated, and the coefficient is small, but adding these controls has substantially changed the result. We now observe a negative and precisely estimated coefficient: additional sibling increases male test score by 3.4% of a standard deviation and reduces the female score by 2.1% of a standard deviation (5.5-3.4). The other important difference between rural and urban samples is that the female dummy variable in itself is statistically significantly negative for the rural sample but positive for urban, which means women on average are doing worse in rural areas (18.8% of a standard deviation) but as good as men do in urban areas (-1% of a standard deviation)

Column 5 additionally tests whether the sibling effect on cognitive test scores is mainly due to the effect on educational attainment or the quality of education is also affected within the same schooling year. The results show that controlling for years of schooling largely reduces the sibling effects; however, the coefficients are still statistically significant at 1% level in both samples, although the size has reduced by almost half (-0.095 to -0.037 for the rural sample; -0.055 to -0.032 for the urban sample).

We also estimate the same regressions using math scores as the outcome variable (see Table 2-G4). The results are largely consistent with that of the vocabulary test performance. Not only are females doing worse than their male counterparts in mathematics, but also having an additional sibling is negatively associated with female test performance, while that is not the case for males. Due to the lack of variation in math test score in the urban sample, we are unable to obtain stable results.

The OLS results suggest a strong Quantity-Quality trade-off effect for females but not for males. The empirical literature which tests the Quantity-Quality trade-off has been mixed. Studies have found negative sibling effect on quality as measured by the level of education [Conley and Glauber, 2006; Rosenzweig and Zhang, 2009; Li et al., 2008; Lee, 2008]. Others have also reported non-negative effects of having more siblings on education [Black et al., 2005; Angrist et al., 2005; Haan, 2010; Angrist et al., 2010]. Our finding seems to be closely linked to the son preference culture. When resources are limited, the needs of daughters are likely to be sacrificed.

Two possible explanations come to mind, and both are related to the son preference. First, in traditional Chinese societies, the eldest son typically is to carry on with the family business and responsibilities. They are also often cultivated as the role model for the younger siblings. Thus, it is common for them to receive the most human capital investment from their parents. In a way, the more siblings, the higher the family responsibility and the more the investment from the parents. If this effect is at work, we would be likely to find that for males, the number of younger siblings is more likely to have a positive effect on their educational outcome. The

Table 2-2: OLS results - Vocabulary test

Dependent var.	Panel A: Rural sample				
Vocabulary test z-score	(1)	(2)	(3)	(4)	(5)
Siblings	-0.124***	-0.096***	-0.083***	0.042***	0.020***
	[0.006]	[0.006]	[0.005]	[0.007]	[0.005]
Siblings*Female		-0.056***	-0.060***	-0.095***	-0.037***
		[0.007]	[0.007]	[0.009]	[0.006]
Female	-0.311***	-0.148***	-0.147***	-0.188***	-0.045***
	[0.022]	[0.028]	[0.027]	[0.026]	[0.016]
Parental schooling years				0.050***	0.006***
				[0.002]	[0.002]
Han Chinese				0.120*	0.076**
				[0.064]	[0.034]
Birth order				-0.021***	-0.011**
				[0.006]	[0.004]
Birth order*Female				0.046***	0.018***
				[0.009]	[0.006]
Age				-0.002	0.006**
				[0.005]	[0.003]
Age squared				-0.000***	-0.000***
				[0.000]	[0.000]
Years of schooling					0.133***
					[0.002]
Birth county dummies	No	No	Yes	Yes	Yes
Birth county dummies*time	No	No	No	Yes	Yes
Observations	21,641	21,641	21,641	21,641	21,641
R-squared	0.100	0.103	0.232	0.419	0.688
Joint significance of Sibling and Sibling*Female (P value)		0.000	0.000	0.000	0.000
Dependent var.	Panel B: Urban sample				
Vocabulary test z-score	(1)	(2)	(3)	(4)	(5)
Siblings	-0.098***	-0.091***	-0.079***	0.034***	0.031***
	[0.007]	[0.008]	[0.009]	[0.012]	[0.009]
Siblings*Female		-0.012	-0.014	-0.055***	-0.032***
		[0.009]	[0.010]	[0.014]	[0.012]
Female	0.006	0.032	0.023	-0.010	-0.023
	[0.015]	[0.023]	[0.023]	[0.029]	[0.028]
Parental schooling years				0.035***	0.008***
				[0.003]	[0.002]
Han Chinese				0.061	0.011
				[0.065]	[0.044]
Birth order				-0.031**	-0.030**
				[0.014]	[0.012]
Birth order*Female				0.049***	0.041***
				[0.016]	[0.013]
Age				0.447***	0.318***
				[0.017]	[0.014]
Age squared				-0.000***	0.000
				[0.000]	[0.000]
Years of schooling					0.103***
					[0.003]
Birth county dummies	No	No	Yes	Yes	Yes
Birth county dummies*time	No	No	No	Yes	Yes
Observations	4,058	4,058	4,058	4,058	4,058
R-squared	0.091	0.091	0.183	0.335	0.543
Joint significance of Sibling and Sibling*Female (P value)		0.000	0.000	0.001	0.002

Note:
1) Robust Standard Errors (SEs) are presented in parentheses;
2) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
3) Test year and county information source indicators are included in column (3)-(5);
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

second possible explanation is related to the sibling spillover effect. Elder female siblings often take the responsibility of childcare [Qureshi, 2018], and female siblings are more likely to take care of other household chores in general. If this is the case, male siblings may have more time to work on their schoolwork, and their parents may also have more time to be with them. Other things being equal, more female sibling may increase male siblings' time invested in their education and improve the quality of their education investment. For this effect to work, we would observe the number of older and younger sisters should exert a positive effect on test scores of their male siblings. We test these conjectures directly by separate the sibling variables into male and female siblings as well as younger and older siblings. The results are reported in Table 2-G3. The table shows that for males, there are strong positive younger sibling effects for both brothers and sisters, indicating the role model effect. In addition, only older sisters have a positive sibling effect but not brothers. This seems to suggest the female sibling spillover effect is also at work. Interestingly, the positive effect is the largest for younger sisters. Perhaps here, both role-model effect and sibling-spillover effects are at work.

Another interesting effect observed is the positive female level effect for the urban sample. As discussed in the background section, the OCP was strictly enforced in urban areas, and hence, since the introduction of the family planning policies in the early 70s, especially after the introduction of the OCP, most urban families only had one child. Thus, despite the prevalence of son preference in the society, when the policy forced upon individuals to have only one child and the child happened to be a girl, it appears that parents poured their resources into the girl, and as a result, girls in urban China are catching up with boys with regard to their school performance.

5.2 IV results

As discussed earlier, the OLS estimates of effects of siblings on the individual standardised cognitive test score are inconsistent due to potential omitted variable bias. We use the instrumental variable approach with the main instrument - Birth Control Rate (BCR) at the county level - and the second instrument - Policy Fertility Index (PFI) as a sensitivity check. Table 2-G14 to Table 2-G15 provide the first and second stage results additionally using PFI. Note that as we are using these variables as instruments for the number of siblings, we generate a seven-year average measure (3 year before and after the birth year of the individual) to capture the potential policy effect on the number of siblings.

The results from the first stage regressions are reported in Table 2-3. Since we have two endogenous variables (siblings and the interaction term between siblings and the female dummy variable), we also use the interaction term of each of our IVs with the female dummy variable to form the second IV. The first two columns report results for the rural sample. As can be seen in the table, BCR is a strong IV for the rural sample. The Kleibergen-Paap Wald rk F statistics for BCR and its interaction term is 155, and Table 2-G14 shows that for the PFI is

18.74.[25] For the urban sample, both the BCR and PFI are rather weak, with the F-statistics being less than 4. We, therefore, will be cautious when discussing IV results for the urban sample.

The direction of the relationship between the IVs and the number of sibling variable also makes sense. We find that the BCR, which measures the intensity of the birth control policy, is negatively correlated with the number of siblings. Table 2-G15 shows that the PFI, which indicates the number of children allowed, is positively correlated with the number of siblings for the rural sample. We also use 3 or 5 year average of the BCR as the instruments, and the results only changed marginally.[26]

Table 2-3: First stage results

	Rural sample		Urban sample	
	(1)	(2)	(3)	(4)
Dependent var.	Siblings	Siblings*Female	Siblings	Siblings*Female
Birth Control Rate (%)	-0.013***	0.002***	-0.006***	0.006***
	[0.001]	[0.001]	[0.002]	[0.001]
Birth Control Rate (%)*Female	0.002***	-0.017***	0.001	-0.019***
	[0.000]	[0.001]	[0.001]	[0.001]
Controls	Yes	Yes	Yes	Yes
Observations	21,641	21,641	4,058	4,058
Kleibergen-Paap Wald rk F statistics	155.21		3.43	

Note:
1) Robust Standard Errors (SEs) are presented in parentheses;
2) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
3) Controls include birth county (or prefecture) fixed effects, age, age square term, birth order, an interaction of birth order with female variable, Han Chinese indicator, parental mean schooling years, county information source indicators, test year indicators, and an interaction between time linear trend variable*birth region fixed effect;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

Table 2-4 presents the results from the IV estimation using the standardised Chinese vocabulary test score as the dependent variable. The left and right panels show the results for rural and urban samples, respectively. The first column of each panel reports the OLS results with the full controls. Column 2 of each panel presents the results using BCR as the IV, which is our preferred estimation. The results using PFI as IV is available in Table 2-G14.

The results in Table 2-4 indicate that the number of siblings and its interaction term with the female dummy have consistent patterns between the OLS and IV estimates. Having an additional sibling increases male test scores while it reduces the female test scores. The magnitude of the interaction term, however, is much larger for the IV estimations than that for the OLS estimation. Having one more sibling increases male test score by 12.6% of a SD and reduces female test scores by 20.6% (0.126-0.332) of a SD. Thus, it widens the gender gap in test scores by 33.2% of a SD, which is approximately three times larger than that estimated from

[25] Stock and Yogo (2005) suggest that when two instruments are used, the critical value of the F test being strong is 19.93 when standard error is not clustered and not correlated for heteroskedasticity.
[26] See Table 2-G8.

Table 2-4: IV results - Vocabulary test

Dependent var.	Rural sample		Urban sample	
Vocabulary test z-score	(OLS)	(IV)	(OLS)	(IV)
Siblings	0.042***	0.126***	0.034***	0.160
	[0.007]	[0.037]	[0.012]	[0.152]
Siblings*Female	-0.095***	-0.332***	-0.055***	-0.098***
	[0.009]	[0.017]	[0.014]	[0.027]
Female	-0.188***	0.060*	-0.010	-0.017
	[0.026]	[0.035]	[0.029]	[0.032]
Parental schooling years	0.050***	0.051***	0.035***	0.036***
	[0.002]	[0.003]	[0.003]	[0.003]
Han Chinese	0.120*	0.119*	0.061	0.063
	[0.064]	[0.064]	[0.065]	[0.067]
Birth order	-0.021***	-0.091***	-0.031**	-0.131
	[0.006]	[0.025]	[0.014]	[0.116]
Birth order*Female	0.046***	0.227***	0.049***	0.088***
	[0.009]	[0.013]	[0.016]	[0.027]
Age	-0.002	0.002	0.447***	0.495***
	[0.005]	[0.007]	[0.017]	[0.073]
Age squared	-0.000***	-0.000***	-0.000***	-0.000*
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	21,641	21,641	4,058	4,058
Kleibergen-Paap Wald rk F statistics		155.21		3.43
P value for Joint significance test				
Sibling and Sibling*Female	0.000	0.000	0.001	0.001

Note:
1) The outcome variables are test z-scores with zero mean and a standard deviation;
2) Robust Standard Errors (SEs) are presented in parentheses;
3) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
4) Control variables include the same set of covariates in Table 3;
5) The instrument is 7-year average Birth Control Rate;
6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

the OLS regression. This may indicate that the OLS estimates suffer from an upward bias. This pattern of having a larger IV estimate relative to the OLS estimate is commonly found in the existing literature, which studies the effect of the number of siblings on children's education or health outcomes [Liu, 2014; Li et al., 2008; Li and Zhang, 2017].

The effect estimated from two different IVs is similar for the interaction term but different for the number-of-siblings level term. The PFI results for the number-of-siblings level term is positive, but the coefficient is minimal and not statistically significant.

For the urban sample (the right panel of Table 2-4) we observe the same pattern, but the magnitudes of the estimates are much smaller. For the urban sample, an additional sibling increases male test score by 16% of a SD (not precisely estimated) and female test score by 6.2% (0.160-0.098) of a SD. In other words, having an additional sibling widens the gender gap in test scores for the urban sample by 9.8% of a SD. The results obtained from using the PFI as the IV are not reliable because the IVs are too weak.

We also estimate the regressions further controlling for the years of schooling, and the results are reported in Table 2-G10. Controlling for years of schooling, we still observe an additional number of sibling effect on test scores, suggesting that having additional siblings reduces female cognitive skills even for those with the same level of schooling.

Table 2-G5 shows that there is also a differential sibling effect on male and female math skills. An additional sibling widens the gender gap in math scores by 23.6% of a standard deviation among rural people using BCR as the IV. However, the results for the math test score are not reliable to some extent. We observe a positive and huge effect of the number of siblings in column 4, and perhaps the spikes in the math test score distribution played a role here.

6 Channels

In this section, we investigate the channels through which the number of siblings played a role in individual test scores.

6.1 Son preference

Although we observe that having an additional sibling is advantageous for boys and disadvantageous for girls and we have conjectured that this could be due to the prevalence of son preference, we have not provided evidence to link these two phenomena. In this sub-section, we try to find such a link.

To do so, we first need to measure the degree of son preference prevalence. The Survey of Women's Status in Contemporary China (SWSC) conducted by the Institute of Population Studies of the Chinese Academy of Social Sciences in 1991 has some interesting questions regarding son preference. The survey asks the respondents, 'If only one child is allowed, do you prefer a son or a daughter?' and each respondent could choose one answer among (1) a

son, (2) a daughter, or (3) indifferent. We use the proportion of women who chose 'a son' over the other two options at the provincial level, by rural and urban separately, to measure son preference level. All the respondents were born between 1926 and 1973.[27]

Figure 2-7: Son preference and sibling effects on gender gap in cognitive skills

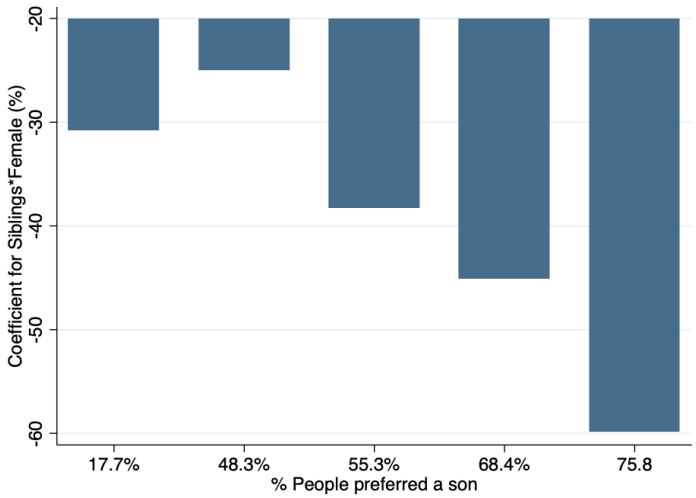


Table 2-G6 presents the proportion of the sampled individuals who chose a son over the other two options (a daughter or indifference) by province and by rural and urban. The table shows that 55.7% of the respondents living in rural and 21.6% in urban regions have a son-preference. The variations within rural sample across provinces are quite large, varying from 17.1% in Shanghai to over 75% in Guangdong and Ningxia provinces. The variation among urban people across provinces range from 18.4% in Shanghai to 34% in Guangdong.

To confirm that our finding of the gender gap on sibling effect is due to the level of son preference, we estimate our main model (column 2 in Table 2-4) by province and plot the coefficient of the interaction term between the number of sibling variable and female dummy variable against the proportion of individuals who have son preference in Figure 2-7 for rural population.[28] It can be seen that the gender gap estimated using the CFPS data is indeed correlated with the share of people with son preference at the provincial level. Provinces with the larger gender gap in the sibling effect also have a higher proportion of people with son preference.[29]

We further check whether our results are significantly associated with traditional beliefs in a son's role in the patrilineal culture. As discussed earlier, Confucian values may lead parents to value sons more than daughters because sons alone can continue family lineage and perform ancestor worships. Also, patrilocality culture which provides that a woman is

[27]The SWSC was conducted in the following provinces: Shanghai, Shanxi, Shandong, Guangdong, Jilin, and Ningxia.

[28]Since the CFPS does not include Ningxia province in its sample, the comparison here only comprises five provinces for which we have both CFPS and SWSC data. The sample size for the urban sample of Shandong, Shaanxi, and Jilin in the CFPS is 28, 71 and 148, respectively. We are unable to obtain sensible estimates for these provinces using their urban survey.

[29]Table 2-G7 provides the estimated results.

exported from her birth family to her husband's family upon marriage may also be the cause of son preference. If that is the case, parents will prioritise sons because sons are more likely to live with them and provide old-age support for them. In fact, a Chinese proverb goes that raising a daughter is ploughing someone else's field [Jayachandran, 2015].

The CFPS asked relevant questions in 2014. The first question is 'Women should give birth to at least one boy to continue family lineage' which captures how much a respondent values the patrilineal system. The second question is 'A man should live with his parents after marriage' which reflects a respondent's belief in patrilocality. The next four questions ask beliefs about gender roles in general. The questions are 'Men should focus on career, while women should focus on family', 'Marrying well is more important for women than doing well', 'Women should have at least one child', and 'Men should do half of the housework'. Each respondent should choose one option among (1) strongly agree, (2) disagree, (3) neutral, (4) agree and (5) strongly agree. We use the proportion of individuals who choose either (4) agree or (5) strongly agree over the other three options at the provincial level by rural and urban separately.[30]

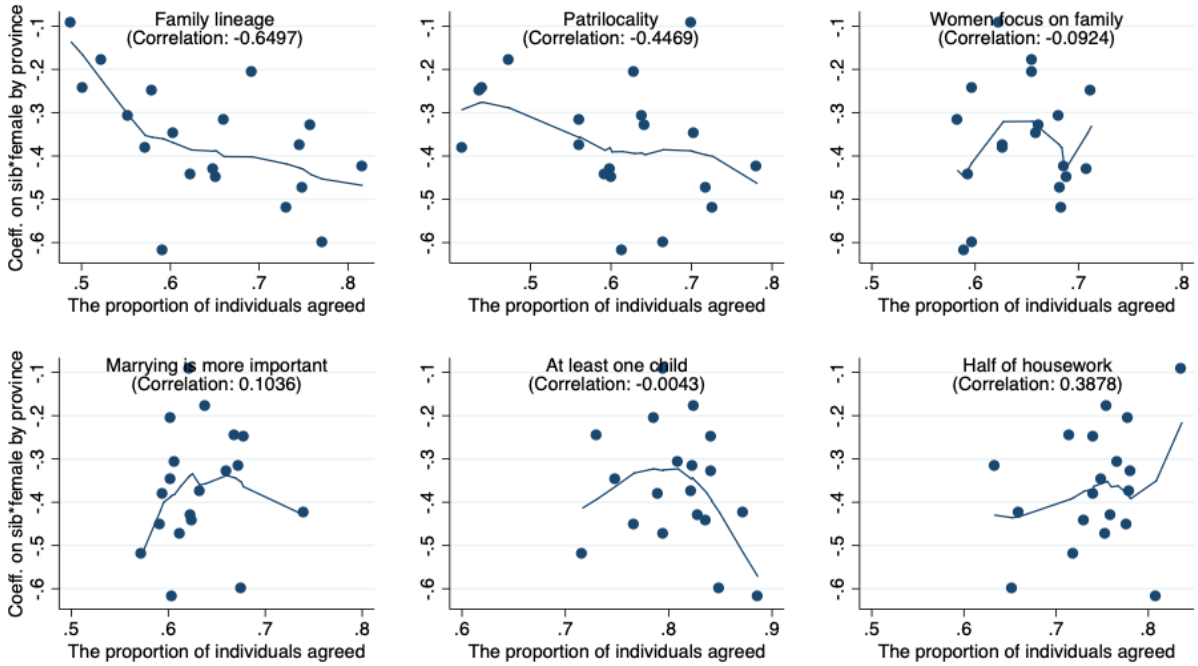
We estimate our main model (column 2 in Table 2-4) again by province and plot the coefficient of the interaction term between the number of sibling variable and female dummy variable against the proportion of individuals who agreed with each statement in Figure 2-8 for the rural population. Originally 25 provinces were surveyed in the CFPS in 2014 but 6 of them had less than 400 individual observations for the main estimation, which means each cell by birth year and gender has less than 5 observations (400 individuals for 41 birth years and two gender groups) so we exclude those 6 cities or provinces.[31]

The first graph in the first row (Q1) indicates that the magnitude of female disadvantage in an additional sibling has a strong negative correlation with the proportion of individuals who agreed that women should give birth to at least a son to continue the family lineage. In other words, the female disadvantage is more pronounced in regions with stronger patrilineal norms. When more than 80% of people agreed with the statement, an additional sibling widens the gender gap in cognitive skills by about 40% of a standard deviation, while it widens only less than 10% of a standard deviation in a province where less than 50% of people agreed. The correlation coefficient is -0.6497, and when we regress the provincial coefficient of the interaction term (values on y-axis) on the proportion of individuals agreed (values on x-axis), the R-squared is 0.4221 as presented in Table 2-G11. The second graph in the first row (Q2) also shows that sibling effects on the gender gap are larger in provinces where people are more likely to believe that a man should live with his parents after marriage. The correlation is weaker compared to the patrilineal norm case in Question 1, and also when we regress

[30] Among the 37,147 individuals who were surveyed in 2014, 84.8% (31,510 respondents) answered Question 1, 84.8% (31,506 respondents) answered Question 2, 84.5% (31,405 respondents) answered Question 3, 84.6% (31434 respondents) answered Question 4, 84.8% (31487 respondents) answered Question 5, and 84.8% (31515 respondents) answered Question 6. The respondents were born between 1912 and 1998.

[31] It includes Beijing (N=44), Tianjin (N=112), Zhejiang (N=390), Fujian (N=238), Hubei (N=340), and Chongqing (N=199).

Figure 2-8: Son preference and sibling effects on gender gap in cognitive skills



Q1: Women should give birth to at least one boy to continue family lineage
Q2: A man should live with his parents after marriage
Q3: Men should focus on career, while women should focus on family
Q4: Marrying well is more important for women than doing well
Q5: Women should have at least one child
Q6: Men should do half of the housework
* Correlation indicates the correlation coefficient between values on Y axis and X axis

provincial coefficients (values on y-axis) on the proportion of individuals agreed (values on the x-axis), the R-squared is 0.1997 which is less than half of 0.4221 in the first question. This suggests that the traditional beliefs that a son only can continue family lineage seem to play a more important role in terms of female disadvantage in big families than patrilocality culture.

The rest of the figures show weaker relationships compared to the first and second figures. Interestingly, the correlation with a belief 'women should have at least one child' is almost zero, while that is 0.6497 when the statement includes 'a son'.

In summary, we find that female disadvantage in having additional siblings is more pronounced in regions with (1) a higher prevalence of son preference and (2) stronger beliefs in a son's role as a successor in Confucian traditions.

6.2 Limited resources

According to Becker's Q-Q model, the negative relationship between child quantity and quality exists due to household budget constraints [Becker and Lewis, 1973]. With son preference, we expect female disadvantage in having siblings could also be larger in families with tighter budget constraints compared to families who are well off. This is because parents who do not have enough resources to educate all their children well could be more likely to disadvantage

daughters to keep enough resources for boys than parents who are well off. In this subsection, we test if this is the case.

To do so, we need to measure household financial conditions in one's early childhood; however, we do not have such household income information in early childhood. Instead, we use parental mean schooling year as a proxy for household income. We divide our main samples into two groups. The first group consists of individuals whose parental mean schooling years are less than or equal to the average of parental mean schooling years among their birth year group. The second group consists of those whose parental mean schooling years are higher than the average among their birth year group. We calculate the average for each birth year group for the rural sample and urban sample separately.

We run our main estimation model (column 1 and 2 in Table 2-4) for each group separately in Table 2-5. Additional adverse sibling effects on the gender gap in cognitive skills are statistically significant for both parental education groups in the rural sample, while it only happens to the lower parental education group in the urban sample. This evidence overall supports Becker's theory considering that our findings come from individuals whose parents are educated for less than 6 years on average (0.97 years for the lower and 5.96 years for the upper group in the rural sample, and 3.32 years for the lower group and 9.31 years for the upper group in the urban sample). Note that parental education level may also reflect parental son preference level not only in childhood environments; this requires careful interpretation of the results.

Table 2-5: Mechanism analysis - Limited resources

	Rural sample				Urban sample			
Dependent var.	**Pedu≤Mean**		**Pedu>Mean**		**Pedu≤Mean**		**Pedu>Mean**	
Vocabulary z-score	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Siblings	0.030***	0.154***	0.054***	0.091*	0.053***	0.408**	-0.003	-0.429
	[0.009]	[0.054]	[0.011]	[0.050]	[0.019]	[0.204]	[0.024]	[0.750]
Sibling*Female	-0.074***	-0.329***	-0.119***	-0.314***	-0.074***	-0.156***	-0.025	-0.039
	[0.011]	[0.025]	[0.012]	[0.024]	[0.019]	[0.054]	[0.022]	[0.047]
Female	-0.324***	-0.040	-0.038	0.162***	-0.054	-0.060	0.028	0.063
	[0.034]	[0.050]	[0.029]	[0.041]	[0.041]	[0.068]	[0.037]	[0.060]
Observations	12,440	12,440	9,201	9,201	2,039	2,039	2,019	2,019
Kleibergen-Paap Wald rk F stat		90.72		93.24		4.22		0.55
P value for Joint significance test								
Sibling and Sibling*Female	0.000	0.000	0.000	0.000	0.001	0.000	0.237	0.703
Average parental schooling years	0.97 years		5.96 years		3.32 years		9.31 years	
Average household income per capita	6640 yuan		9623 yuan		14488 yuan		21761 yuan	

Note:
1) Mean parental schooling years is the average of parental schooling years for each birth cohort among rural individuals and urban individuals, respectively;
2) Within each group, the average parental schooling years is 0.97 for the lower parental education group and 5.96 years for the upper group in the rural sample. It is 3.32 for the lower group and 9.31 years for the upper group in the urban sample;
3) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
4) Robust Standard Errors (SEs) are presented in parentheses;
5) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
6) Controls include the same set of covariates in the main model in Table 4 unless otherwise indicated;
7) The 7-year average Birth control rate is used as the instrument;
8) Average household income per capita is as of 2010 and includes only households that were surveyed in 2010;
9) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

7 Robustness check

7.1 Confounding factors

China had experienced socioeconomic environment changes to a great extent over the past decades. Those changes may confound sibling effects on cognitive skills because individuals in our sample would be exposed to such changes at the different development stages to different extents, and we are unable to control for all these factors.

As a solution, our model includes age, age squared term and a linear time trend variable for each county by assuming that either linear or quadratic time trends would capture those environmental change impacts on cognitive skills. However, the relationship may also be nonlinear and even discontinuous. For example, during the Cultural Revolution period from 1966 to 1976, those born between 1953 and 1963 missed more than 1 year of primary school education due to school closure, although those born before 1953 or after 1963 did not experience it.

Another solution could be controlling for birth-year fixed effects. However, including birth-year dummy variables weakens our instruments because the birth control policy intensity mainly varies by time due to the government policy scheme change over time. This means that controlling for birth-year fixed effects may take both effects of environmental and birth control policy changes at the same time. That is, the specification controlling for birth-year fixed effects could be one of the solutions; however, it is not a better solution compared to the main specification because it washes away the exogenous shock on fertility from the instruments.

Column 2 in Table 2-6 replaces age and age squared term in our main model (column 2 in Table 2-4) with birth-year fixed effects and its interactions with female dummy. Above all, we do not see a major change in the coefficient of our interest. An additional sibling widens the gender gap in cognitive skills by 33.0% of a standard deviation in the rural sample, while the effect was 33.2% of a standard deviation in the main results shown in column 1. In the urban sample, the magnitude of the coefficient is double (18.1%) in column 2 after controlling for birth year fixed effects than 9.8% in column 1. Kleibergen-Paap Wald rk F statistics is 36.89 for the rural sample and 1.08 for urban sample, which is about five times smaller than those for the main results in column 1.

As mentioned earlier in Section 2 (Background) individuals born before 1965 could be exposed to the government gender equality campaign to a great extent in their childhood, which could affect the gender gap in cognitive skills among them. Column 3 in Table 2-6 excludes individuals born before 1965 and shows that the coefficients on the interaction term of our interest are negative and statistically significant at 1% level in the rural sample. The coefficient on the interaction term becomes close to zero (0.005), and the negative coefficient on female dummy is 10 times larger in an absolute term among urban individuals. However, Kleibergen-Paap Wald rk F statistics is only 1.08, which may not be a reliable result.

Table 2-6: Robustness check

	Rural sample			
	(1)	(2)	(3)	(4)
Dependent var.	Main	Birth year	Born	Excl.
Vocabulary z-score	result	fixed effect	after 1965	Migrants
Siblings	0.126***	-0.239	-0.095*	0.142***
	[0.037]	[0.216]	[0.053]	[0.044]
Siblings*Female	-0.332***	-0.330***	-0.308***	-0.338***
	[0.017]	[0.045]	[0.060]	[0.020]
Female	0.060*	0.171*	0.565**	0.069*
	[0.035]	[0.092]	[0.261]	[0.041]
Observations	21,641	21,641	13,693	17,513
Kleibergen-Paap Wald rk F statistics	155.21	36.89	18.49	143.93
P value for Joint significance test				
Sibling and Sibling*Female	0.000	0.000	0.000	0.000
	Urban sample			
Dependent var.	Main	Birth year	Born	Excl.
Vocabulary z-score	result	fixed effect	after 1965	Migrants
Siblings	0.160	0.384	0.172	0.085
	[0.152]	[0.551]	[0.168]	[0.192]
Siblings*Female	-0.098***	-0.181*	0.005	-0.100*
	[0.027]	[0.101]	[0.092]	[0.054]
Female	-0.017	-0.096	-0.196	0.043
	[0.032]	[0.094]	[0.322]	[0.038]
Observations	4,058	4,058	2,364	2,395
Kleibergen-Paap Wald rk F statistics	3.43	3.33	1.08	2.62
P value for Joint significance test				
Sibling and Sibling*Female	0.002	0.038	0.292	0.062

Note:
1) Column 1 is the main IV result from Table 4; Column 2 includes birth year dummy variables and its interactions with female dummy instead of age and age squared; Column 3 includes individuals born after 1965 only; Column 4 excludes migrants to the survey counties;
2) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
3) Robust Standard Errors (SEs) are presented in parentheses;
4) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
5) Controls include the same set of covariates in Table 4 unless otherwise indicated;
6) The 7-year average Birth control rate is used as the instrument;
7) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

Another concern is that we use the current residential county information instead of birth county information for 19.07% of rural individuals and 40.98% of urban individuals in our samples. It is because, first, they were born in counties that are not surveyed by the CFPS, so we do not have that county information such as the BCR at the county level. Second, some individuals did not report their birth county information, and we assume that they were born in the current residential counties. The fact that they are currently living where they were not born means they are (possibly) migrants. Including those migrant observations may cause an omitted variable problem in our main estimates for the following reasons. First, we are unable to control for unobservable characteristics of their childhood environments, which are captured in the birth-county fixed effects in our main estimation model. Second, migrant parents may have systematically different child-rearing styles compared to non-migrant parents, which also could not be fully controlled for in our estimation model.

Column 4 only includes individuals who are currently living in the counties where they were born or lived in early childhood at age 3. That is, we exclude migrants who had moved to the current survey counties. In column 4, we rerun the main regression (column 2 in Table 2-4) and find a major change compared to the main results in column 1.

7.2 Cognitive skill measure

We measure individual cognitive skills using the CFPS vocabulary and math test scores in 2010 and 2014. More specifically, we first standardise the raw scores among all the test takers by test year then use the individual average of the standardised scores between 2010 and 2014 for the main outcome variable. In this subsection, we test whether our main results are subject to any measurement errors in terms of individual cognitive skill. First, we check if the main results are sensitive to how to generate the main outcome variable. Second, we use another two measures of cognitive skills. The first measure is individual years of schooling, and the second one is the national college entrance exam score from another data set.

We use (1) the maximum of individual vocabulary test scores between 2010 and 2014 instead of the average, (2) the raw vocabulary scores instead of the standardised vocabulary scores, and 3) the standardised vocabulary test score in 2010 only and 2014 only, separately, and rerun our main estimation model (column 2 in Table 2-4) in Table 2-G12. Both IV and OLS results suggest that our main findings are not sensitive to how to generate the outcome variable.

Second, instead of individual test scores, we use individual years of schooling as the outcome variable in Table 2-G2. An additional sibling widens the gender gap in schooling years by from 0.431 to 1.259 years in the rural sample and from 0.226 to 0.317 years in the urban sample. The estimated coefficients on the interaction between the number of siblings and the female dummy variable are consistently negative and statistically significant in most cases.

Lastly, we additionally use the individual test scores of the National College Entrance

Examination (NCEE) from the Urban Residents Education and Employment Survey (UREES). While the CFPS respondents have neither academic nor financial incentives to achieve high scores in the CFPS tests, the NCEE is the real high-stake exam.[32]

The NCEE is a centralised examination in China, and students who complete three years of the academic senior high school attend the NCEE in order to apply for universities. The NCEE score mainly determines the admission outcome, which means, again, the NCEE is a real high-stake test in China. The UREES conducted by the National Bureau of Statistics (NBS) of China provides individual self-reported NCEE total scores (see Fan et al. [2018] for more information). The NBS randomly selected 10,000 households in 2004 from 12 provinces in China and urban areas, and asked each respondent for information regarding their last NCEE experience (e.g. test year, test province, test type, and the total score in addition to their family size information). This sample consists of 4,753 individuals born between 1951 and 1987 who attended the general college entrance exam and whose household registration is urban. Also, the sample excludes those who attended the NCEE before 1977 because the exam was officially suspended during the Cultural Revolution (1966-1976) and so the exam attendants are highly likely to be selective.[33]

Table 2-G13 provides the estimation results. We add test year fixed effects and test province fixed effects into our main model (column 2 in Table 5). Note that the UREES surveyed only urban residents. Column 1 shows one more sibling widens the gender gap in the NCEE standardised total score by 6.9% of a standard deviation, and the coefficient on the interaction term is statistically significant at 1% level. Note that we have the BCR information for the prefectures that are surveyed by the CFPS. Columns 2 and 3 includes individuals living in the prefecture surveyed by both the CFPS and UREES.[34] Female disadvantage in having an additional sibling ranges from 5.6% to 12.5% of a standard deviation in column 1-3, which does not significantly deviate from 8.7% of a standard deviation in our main result using the CFPS vocabulary test score.

7.3 Birth order effect

Black et al. [2005] shows that a negative coefficient of family size on individual educational attainment becomes negligible when birth order is additionally controlled for in the Norwegian case. Since birth order is highly correlated with the number of siblings, without controlling for birth order, we are unable to disentangle family size effect from birth order effect. As an example, a firstborn is always likely to be in a small family than other children born later. In our sample, the correlation coefficient between the two variables is 0.6554 in rural and 0.7448

[32] A high correlation between individual schooling years and the CFPS test scores suggests that the CFPS test scores are good proxies for individual cognitive skills. (vocabulary test - rural sample: 0.7868; urban sample: 0.6867; math tests - rural sample: 0.8862; urban sample: 0.8487).

[33] See Appendix H for more details about sample construction and survey province.

[34] Kleibergen-Paap Wald rk F statistic suggests the instrument is very weak (1.38), and it is partially due to that we include test year dummy variables because the Birth control policy intensity varies mainly by time. Excluding those dummy variables, F statistic is 7.270.

in the urban sample.

For this reason we control for birth order and its interaction with female dummy variable in the main model. In addition, we here estimate the main model (column 2 in Table 2-4) using firstborn individuals and non-firstborn individuals separately in Table 2-7. Column 2 and 5 show, among firstborn people, the coefficients on the interaction of sibling number variable with female dummy are statistically significant at 1% level and consistently negative. That is, the female disadvantage of having one more sibling exists within the same birth order group as well.

The magnitude of the negative coefficient on the interaction between sibling number and the female dummy is larger among firstborn individuals (in column 2 and 4) than among non-firstborn individuals (in column 3 and 5) which is consistent with the patterns in Table 2-6 and Table 2-G9. One more sibling widens the gender gap by 34% of a standard deviation among firstborns, as compared with 22.9% of a standard deviation among individuals who are not firstborn. The difference is smaller in the urban sample, which is only 0.3% of a standard deviation. Since firstborns do not have older siblings, the number of siblings is equal to the number of younger siblings for them. That is to say, younger siblings seem to contribute the most to widening the gender gap, especially in rural regions where son preference is prevalent. It is consistent with the results in Table 2-G3 that female disadvantage in having siblings mainly comes from younger sibling effects.

Table 2-7: First born only

	Rural sample			Urban sample		
Dependent var. Vocabulary z-score	All	Firstborn	Not Firstborn	All	Firstborn	Not Firstborn
Siblings	0.126***	0.097	0.134***	0.160	0.095	0.013
	(0.037)	(0.065)	(0.041)	(0.152)	(0.181)	(0.197)
Siblings*Female	-0.332***	-0.340***	-0.321***	-0.098***	-0.089***	-0.129***
	(0.017)	(0.026)	(0.022)	(0.027)	(0.029)	(0.049)
Female	0.060*	0.358***	-0.074	-0.017	0.068**	0.021
	(0.035)	(0.052)	(0.045)	(0.032)	(0.033)	(0.063)
Observations	21641	7404	14237	4058	1949	2109
Adjusted R^2	0.374	0.379	0.359	0.267	0.332	0.228
Kleibergen-Paap F stat.	155.21	40.39	110.06	3.43	2.49	9.37
P value for Joint significance test						
Sibling and Sibling*Female	0.000	0.000	0.000	0.002	0.007	0.039

Note:
1) Controls include the same set of covariates in Table 4 unless otherwise indicated;
2) The 7-year average Birth control rate is used as the instrument;
3) Robust Standard Errors (SEs) are presented in parentheses;
4) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
5) County and prefecture fixed effects are included for the rural and urban samples, respectively;
6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Conclusion

This study explores the relationship between fertility and the gender gap in cognitive skills. Over the past decades, gender gaps in educational and labour market outcomes which favour men have been narrowed. At the same time, fertility has significantly declined over time. We link these two phenomena. The idea is based on Becker's Quantity-Quality trade-off, which shows a negative relationship between family size and the average quality of child due to parental budget constraints. Our hypothesis is, in the presence of son preference in a family, the negative relationship is more distinct for daughters, especially when the family faces tight budget constraints. That is, the negative effect of an additional sibling is more significant for daughters, and it widens the gender gap in child quality by worsening female quality more than it does for male quality.

Existing studies on China's One Child Policy have discovered the family size effect on individual quality. However, as discussed in Zhang [2017], a lesson drawn from their findings for other developing countries may be limited, especially considering that 1) the fertility shock marginally decreased family size from 2 to 1 or 1.5 to 1, and 2) the policy initiated almost at the same time as the market-oriented economic reform. In this sense, Zhang (2017) states that "a developing country may want to consider an aggressive but less-coercive family planning program such as the policy implemented in China in the 1970s.".

This is the first study that tests the Q-Q trade-off theory by exploiting the fertility policy shock, which decreased the national fertility rate from 6 children to 2 children per woman since the early 1970s, to our best knowledge.

Using the exogenous shock on fertility as an instrument, we find strong and statistically sibling effects on the gender gap in cognitive skills in both urban and rural areas. An additional sibling widens the gender gap in the vocabulary test scores by 5.1% to 26.5% of a standard deviation. The adverse effects are substantial in regions with a high prevalence of son preference and households with tight budget constraints.

An important contribution of this study is the external validity of our results. First, as mentioned earlier, we exploit the first exogenous shock on fertility in China since the early 1970s. Second, the main sample is nationally representative because it includes individuals from 25 provinces in China, which covers most of the provinces. Third, we measure cognitive skills in adulthood, in addition to the completed years of schooling. This measure may more directly relate to individual educational and labour market outcomes in adulthood. Additionally, this paper closely looks into the differential sibling effects between genders and provide suggestive evidence as to what cultural norms or beliefs could drive female disadvantage in a family.

Explaining Trends in Adult Height in China: 1950 to 1990

1 Introduction

The average height of individuals is an important marker of the health conditions that they faced during childhood, especially in the first few years of life. Although genetic variation accounts for around 80 percent of the variation in height between individuals, early childhood conditions have effects that can be identified both within and between population groups [Silventoinen, 2003; Steckel, 2009]. Environmental conditions are transmitted through two main channels: nutrition and disease, both of which strongly influence growth during infancy and early childhood. Inadequate nutrition slows growth and may result in stunting which cannot be fully recovered in adolescence. Exposure to disease diverts the body's energy to fighting repeated infections, rather than building bone and muscle [Beard and Blaser, 2002].

A variety of historical studies have analysed adult heights to explain historical trends and associations between adult height and conditions in the household and the locality around the time of birth. In Western Europe, the average height of adult male cohorts increased by about 1 cm per decade from the 1870s to the 1970s. Many of the earlier studies on factors affecting height are outlined in surveys by Steckel [1995, 2009]. Using aggregate time-series data for Spain from 1850 to 1958, Maria-Dolores and Martinez-Carrion [2011] finds that height was positively influenced by GDP per capita and the share of health expenditure in total expenditure, and negatively correlated with the mortality rate and the price of consumption goods. Using panel data on the average heights of adult males in 15 European countries from the 1870s to the 1970s, Hatton [2014] finds that both living standards and the disease environment contributed significantly to the long-run increase in height. Other factors such as the decline in family size (the quality-quantity tradeoff), the increase in parental education, and the expansion of welfare services each made a modest contribution. Studies of individual-level adult height support these findings. For example, among prisoners in nineteenth century Bavaria, Baten and Murray [2000] find that women's heights were more sensitive than men's to economic, nutritional and disease conditions. Using a cross-section of British servicemen measured during the First World War, Bailey et al. [2016, 2018] find that the socioeconomic status of the household head, family size, and household composition all influenced adult

height, but that the local environment, particularly industrial pollution, mattered even more.

In the past sixty or so years, China has experienced turbulent and significant socioeconomic transformations. The industrialisation process since the early 1950s was accompanied by the following major economic and social upheavals: the Great Famine (1959-1961), the Cultural Revolution (1966-1976), the introduction of loose family planning policies (1973-1978) and the One Child Policy (OCP, 1979 onwards), and the market-oriented economic reforms (1979 onwards). These events generated shocks in many factors which are regarded as closely associated with height: nutrition intake, education, fertility, and household income.

A few existing studies have examined the effects of a specific event or policy on adult height of Chinese people. A commonly studied event on height in China has been the Great Chinese Famine ([Chen and Zhou, 2007; Meng and Qian, 2009; Wang et al., 2010; Gørgens et al., 2012; Li and An, 2015]). Gørgens et al. [2012], for example, find that those exposed to the famine when under five years of age were stunted by 1-2 cm but that this was exactly offset by positive selection among the survivors. Studies of the effect of the Cultural Revolution, which reduced the education level of the parental generation, find that it reduced the height of their children [Chen, 2010]. Many studies also examined the OCP effect, that reduced fertility, and increased children's height-for-age z-scores [Liu, 2014; Kubo and Chaudhuri, 2017].

Studies that examine the change of adult height and its contributing factors in a longer time horizon for China are rare owing to lack of data for the early period. Morgan [2000] find that after a modest advance from the 1920s to the 1950s, heights increased more rapidly in the following decades. Consistently, Bi and Ji [2005] find that between 1956 and 2000, the mean heights of 18-years-old (most closely reflect the adult height)[1] for male and female students increased from 164.9 cm to 172.6 cm and from 155.9 cm to 160.6 cm, respectively. These correspond to 1.75 cm and 1.07 cm per decade for males and females, respectively. Other studies using survey data also find heights advanced over a similar period (See [Schwekendiek and Baten, 2019] for example). None of these studies, however, has examined the extent to which the economic and health environmental conditions at the crucial developmental stages of these cohorts might have influenced the change in their height over this period.

This paper fills in this gap. We link individuals' adult height with macro-economic and health conditions at their birth county during their childhood crucial developmental ages, together with individual and family characteristics. The purpose of this paper is to answer the following three questions. (1) What are the general trend of height between the 1950s and 1990s and how did it vary between urban and rural regions and by gender? (2) How have changes in economic and social conditions and policies during an individual's crucial height developmental age affected their heights (and health)? (3) How has the family-planning policy, together with son-preference, affected different urban and rural regions and gender differentially?

To do so, we built a unique set of data on county-level economic conditions (industrial

[1] Based on WHO height-for-age chart (see https://www.who.int/growthref/cht_hfa_boys_z_5_19years.pdf?ua=1) the mean height for boys would continue to grow at a lesser rate after 18 years of age.

and agriculture outputs or total outputs), a general health indicator (mortality rate), and local government expenditure on social services between 1950 and 1990. We manually collected these data mainly from hundreds of local gazetteers written by each county local historians. It allows us to align individual's adult height with their birth county local conditions at the time when it was the crucial individual height developmental stage (ages 0-3) for each individual. The information on individuals' adult height was obtained from the China Family Panel Study (CFPS). This survey has two important advantages. One is that it is a nationally representative household survey. The other more important advantage of the CFPS is that it records individuals' childhood information: birth county, household registration (*hukou*) status at age 3, and some basic parental information at the time the person was aged 14. It allows us to link an individual's adult height to the local socioeconomic conditions that he or she faced during early childhood.

We find that the average height growth between 1950 and 1990 for rural and urban males was 1.3 cm and 1.5 cm per decade, respectively. For rural and urban females, the growth was lower than their male counterparts, at 0.7 cm and 1.1 cm per decade, respectively. The height growth was not linear, and the 1980s saw the most significant height growth for all four population groups. The general picture is that height growth was more in urban areas than in rural areas, and more for male than for female population groups. We find that the strong positive time trend on height growth is related to the positive association between height and the increase in local economic circumstances. The positive effects coming from the agricultural and industrial output are stronger for the rural than for the urban sample and more important for the male than the female sample. However, among all local economic and social condition variables which affect height growth, government social spending per capita (spending on health and education) plays the most important role for all samples. The strong economic and social condition effect mainly occurred during the period of economic reforms. We further discovered that the impact of the family planning policy, together with son-preference, have had a negative impact on height of rural population, in particular on rural females.

The rest of this paper is organised as follows. Section 2 provides background of socioeconomic changes in China between 1950 and 1990, and the details of economic reforms and family planning policies made since 1979. Section 3 provides model specifications and estimation issues, and Section 4 introduces the data and the main sample for our analysis. Section 5 presents the main results and potential channels of our main findings. Section 6 provides the robustness check results. Section 7 concludes.

2 Background

In this section we describe socioeconomic and policy changes that are likely to have affected the height of those people born in the 40 years from 1950 to 1990. The modern history of China in this period consists of two major parts: first, a centrally-planned economy since the establishment of the People's Republic of China (PRC) in 1949; and, second, market-oriented

economic reforms since 1978.

2.1 A centrally-planned economy 1949-1977

2.1.1 The establishment of urban healthcare system: 1949-1955

On 21 September 1949, the Communist Party Chairman, Mao Zedong, proclaimed the establishment of the PRC then the Chinese Communist Party (CCP) came to power. At that time, China was suffering from severe poverty and health problems such as malnutrition. The infant mortality rate in 1949 ranged from 12% to 20% among live births in rural regions and almost a third died before the age of five [Liang et al., 1973; Salaff, 1973].[2]

Although the CCP introduced a planned economy upon establishment, during its first few years the free market economy and centrally-planned economy coexisted until 1955 [Bettelheim, 1988]. In rural areas, land reform ensured all farmers possessed a small piece of land. During the early years, the GDP growth rate was relatively high, albeit rising from a very low level [NBS, 1999].[3]

On the establishment of the PRC, the Chinese government introduced a health system, which emphasised equity on the demand side and the central role of the government on supply side [Yip and Hsiao, 2015].[4] Central and local governments owned, funded, and operated all health facilities and clinics in urban areas, and physicians became employees of the state [Yip and Hsiao, 2015].

The urban healthcare system consists of two main schemes: the Government Insurance Scheme, which mainly covered public servants and university staff, and the Labor Insurance Scheme, which covered employees in state-owned enterprises [Xin et al., 2014]. The benefits included comprehensive welfare, such as old-age pensions and maternity benefits, in addition to free medical care [Meng, 2000].

As the prices of all the goods and services were determined by the central government, the cost of provision of public services were kept at a very low level in line with the low wages received by urban workers. It is important to note, though, that this system only covered urban people. For the majority of rural dwellings (accounting for over 80% of the population), their health care, together with all other types of welfare, were to covered by a cooperative system at the village or commune level (tens of villages) to be established after the collectivisation of agriculture in the late 1950s.[Meng, 2000; Xingzhu and Huaijie, 1992].

2.1.2 The planned economy and rural healthcare system: 1956-1958

From 1956, the Chinese government speeded up the nationalisation of all non-agriculture parts of the economy. Enterprises were owned by the state, they produced goods in line with the

[2] According to [Salaff, 1973], there was very little difference in infant mortality between rural and urban China at that time.

[3] The annual growth rate of gross domestic product (GDP) per capita was 8.72 percent in China between 1952 and 1955[NBS, 1999]

[4] See Liang et al. [1973] for more detailed policy information

national plan, and all profits belonged to the state [Wang, 2018]. In rural areas, a collectivisation movement began to encourage individual farmers to pool their land ownership together to form production teams, which was initially based on natural villages. Soon after, the movement pushed production teams to form a larger collective unit (the brigade), while several brigades formed the communes [Lin, 1988]. Under the commune system, all resources were collectively owned and allocated. Output produced by a production team were distributed to the team households first to meet their basic needs. The rest of the output, in excess of basic needs, was bought by the government at prices lower than local market prices, and the revenues were used for reinvestment and providing public services, while remainders were equally distributed across households [Lin, 1988; Meng, 2000]. The national economic plans fully focused on boosting industrialisation at the expense of agriculture [Chow, 1993].

During this period, communes and brigades began to establish their own medical care clinics to provide primary health services, which gradually evolved into the cornerstone of the rural healthcare system known as the Cooperative Medical System (CMS) [Zhu et al., 1989; Yip and Hsiao, 2015]. Some local farmers received basic medical training and became the main medical service providers in rural China [Zhu et al., 1989; Yip and Hsiao, 2015]. The CMS spread quite fast and about 20% of the production brigades in China had established the system by the end of 1958 [Xingzhu and Huaijie, 1992]. As medical services were provided within the communes and brigades, medical professionals were treated like farmers and subject to the same distributional rules. This allowed medical service prices to be kept low [Yip and Hsiao, 2015].

2.1.3 The Great Leap Forward and the Great Famine 1958-1961

Together with the establishment of the communes in rural areas, Chairman Mao proposed a national goal to overtake Great Britain in industrial production within 15 years and launched the Great Leap Forward movement in the spring of 1958. The unrealistic industrialisation target pushed the workforce, including farmers, to abandon agriculture production to focus on primitive steel production which largely resulted in useless outputs. In addition, the aggressive push for rural collectivisation dramatically reduced farmers' incentive for agriculture production. These, together with the inflexible grain procurement system and many other shocks, resulted in the Great Famine [Lin, 1990; Li and Yang, 2005; Gørgens et al., 2012; Meng et al., 2015]. Between 1959 and 1961, malnutrition was widespread, and an estimated 16.5 to 45 million people, mostly living in rural areas, perished from hunger or disease [Ashton et al., 1984; Chen and Zhou, 2007; Li and Yang, 2005; Meng et al., 2015].

2.1.4 Recovery from the Famine 1962-1965

After the crisis, the government brought back its pre-Great Leap Forward system for agricultural production and restored the output level [Meng et al., 2015]. More attempts were made to improve production efficiency. For example, from 1962 the government accelerated the use

of irrigation system and chemical fertiliser which improved production [Lin, 1990]. Total grain output began to increase from 147.5 million tons in 1961 to 160.0 million tons in 1962, an 8.5 percent increase. Mortality rates fell back to the pre-famine trend from 1962.

In this period, the government also increased expenditure on disaster relief considerably. In 1964, 1,211 million yuan was spent on disaster relief, which doubled the amount spent in 1961 [NBS, 1999]. The annual growth of the total social welfare funds was 7.1% between 1962 and 1965 [NBS, 1999]. In addition, this period also saw a substantial increase in the quantity and quality of medical services. During the Famine, the government started to expand medical services and facilities. This was continued during the recovery period, mainly focusing on disease prevention. There were only 50 laboratories for medicine and chemical reagent tests in 1958, but by 1965 a total of 151 laboratories had been established [NBS, 1999]. The number of specialised prevention stations increased by 16% between 1963 and 1964, while the annual growth was only 0.9% between 1960 and 1963 [NBS, 1999].

Although the health services diffused rapidly after the Great Famine, most of the government-provided services were focused in urban centres. Rural health for more than 80% of the population was left to the collectives to administer [Zhang and Unschuld, 2008].

2.1.5 Cultural Revolution 1966-1976 and family planning policies 1973-1978

The recovery did not last long. To preserve the momentum of the communist ideology and restore his power in the CCP, Mao launched the Cultural Revolution in 1966 [MacFarquhar and Schoenhals, 2006; Booth et al., 2018a]. The main target was to purge capitalists inside the Party and traditional 'anti-revolutionary' elements. All schools in urban areas were closed for many years. Urban production and some rural agricultural production was stopped [Booth et al., 2018a]. Such a political turmoil lasted for 11 years and seriously further disrupted economic growth [Borensztein and Ostry, 1996]. The annual growth rate of per capita GDP dropped to negative in the early, chaotic years of the Cultural Revolution, but later production went back to the pre-Famine years.

During this period, government investment in health services was limited and the training of medical personnel in urban areas collapsed as medical schools were closed until 1972 [Dong and R. Phillips, 2008]. However, as the Party was more concerned about inequality between cities and the countryside, rural areas received more attention. During this period, the village and commune-level healthcare services were formally evolved into the CMS. By mid-1970s, almost 90% of the rural population participated in the system [Xingzhu and Huaijie, 1992; Liu, 2004] which provided equal access to inexpensive medical care [Tu, 2016].

To provide sufficient healthcare workers, a nationwide training programme was implemented. Many farmers, who were either middle-school graduates from rural schools or urban youth who were sent down to the countryside during the Cultural Revolution, were given some basic training to become health workers, the so-called 'barefoot doctors' [Hesketh and Wei, 1997; Zhang and Unschuld, 2008; Dong and R. Phillips, 2008; Hipgrave, 2011]. As a result,

the number of rural health workers increased from 1.0 million in 1970 to an estimated peak of 1.8 million in 1977 [Hesketh and Wei, 1997; Hipgrave, 2011].[5] The barefoot doctors promoted basic hygiene, preventive healthcare and family planning, and treated common illnesses in the countryside [Gong and Chao, 1982]. The CMS system is widely regarded as being successful, and between 1949 and 1975, infant mortality dropped from 20% to 4.7% [Liu et al., 1995].

This period also saw the initiation of family planning policies. In the early 1970s, post-famine population growth and the memory of the hunger during the famine had ignited a philosophical debate over the Malthusian Population Trap. Soon after, the policy of "Later, Longer, and Fewer" was introduced at the end of 1973 [Center for Population Studies, 1986; Peng, 1991; Feeney and Wang, 1993; Cai, 2010], which encouraged couples to get married later, have a longer period in between two children, and have fewer children. The policy had a significant impact on the birth rate, especially for the urban population. Over the 1970s, the total fertility rate went down from above five to just around two [Wang, 2011].

2.2 Market-oriented economic reforms and the One Child Policy

2.2.1 Economic reform and health care system evolution

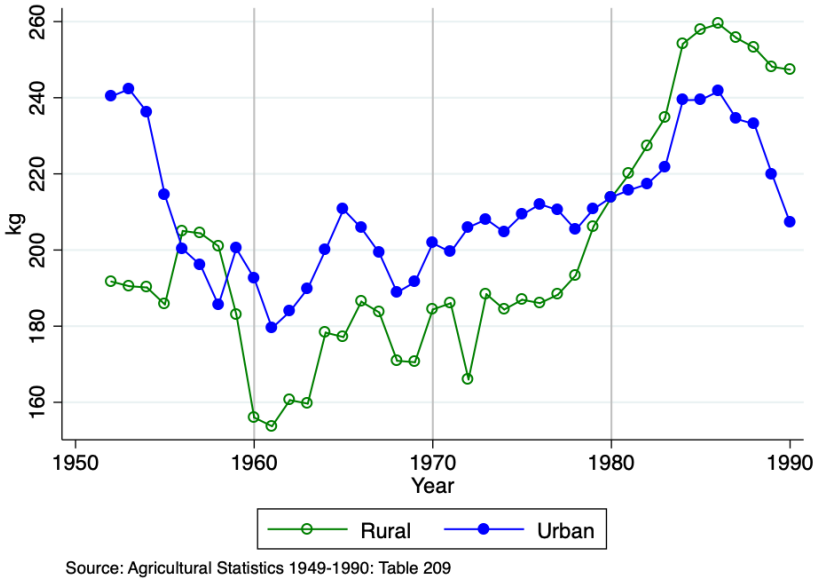
Although efforts were made to improve basic living standards and to reduce inequality during the first 30 years of the People's Republic of China, the lack of economic incentives had hampered economic development. After 30 years of socialist revolution, the system had failed to improve people's living standards substantially. By 1977, 30% of the rural population in China was still living in poverty [Wang, 2018]. In a reaction to poor living conditions in rural areas, small privatisation movements quietly began, whereby farming families cultivated their own contracted land and earned income from what they produced. These activities resulted in bumper harvests [Lin, 1992; Meng, 2012]. The then central government accepted the positive effect of agriculture privatisation and officially introduced this new system - the Household Responsibility System (HRS) - to the whole agriculture sector in China in 1978.

The HRS changed the agricultural production system from collective farming to household-based farming which increased agricultural productivity and, gradually, the income levels of the farmers. Accordingly, the persistent problem of food shortage was resolved [Wang, 2018]. From 1977 to 1984, the total grain output increased from 283 million to 407 million tonnes. The gross value of agricultural output increased by 26% between 1980 and 1983 (at constant prices of 1980) [Lin, 1988]. Figure 3-1 shows the amount of grain consumption per person in rural areas rapidly increased after the introduction of HRS at the end of the 1970s and quickly surpassed that of urban population consumption from 1980.

Although privatisation in agriculture successfully incentivised farmers and significantly increased agriculture productivity and rural income, it destroyed the old collective system - the communes. As the communes dissolved, so too did the social services the system had provided to the rural society. Teachers and healthcare workers in rural areas, who were paid

[5] See [Hu, 2013] for details on sent-down youth during the Cultural Revolution period.

Figure 3-1: Grain consumption per person: 1950-1990



Notes: Grain consumption is calculated on a trade weight (processed or milled) basis. The data used are from Table 209 of Colby et al. [1992]. Their data were gathered from the following yearbooks: 1) China Commerce Yearbook (ZllOngguo Shangye Nianjian) 1988, 2) China State Statistical Bureau, Trade Goods and Materials Statistics Division (Zhongguo Shangye Waijing Tongji Ziliao), 1952-88, and 3) China Commerce Yearbook (Zhongguo Shangye Nianjian), 1991.

by the production teams, brigades, and the communes, now lost their income source and were unable to provide the basic education and health care. In 1978, 82% of all brigades were under the CMS, but by 1983 only 11% brigades remained in the scheme [Huang, 2011; Bloom and Xingyuan, 1997]. With the exception of vaccination and family planning, rural healthcare in the 1980s to the early 2000s relied mostly on private health services [Liu et al., 1995; Blumenthal and Hsiao, 2005].[6]

Furthermore, in the early economic reform period, growth targets were regarded as more important than providing social services. One of the steps taken was to reform the public finance system. The 1994 public finance reform allowed the central government to gain more revenue, but delegated fiscal expenditure responsibilities to sub-national governments [Huang, 2011; Shen et al., 2012]. Consequently, health expenditure began to rely on local government fiscal capacity (Park 1996). Overall, the share of health services financed by the central government fell from 40% in the early 1980s to less than 20% in 1993, and a majority of this funding was spent in urban areas [Huang, 2011].

The urban economy also experienced significant changes. Privatisation shifted a large proportion of workers from state-owned to privately-owned work units. For a long while, the cradle-to-grave social welfare system enjoyed by urban residents in the pre-reform era gradually changed. Public hospitals were no longer completely funded by the state, but increasingly relied on selling services and medicines to pay salaries and bonuses to their staff. However, these changes to the urban sector occurred after the mid-1990s, which is beyond the

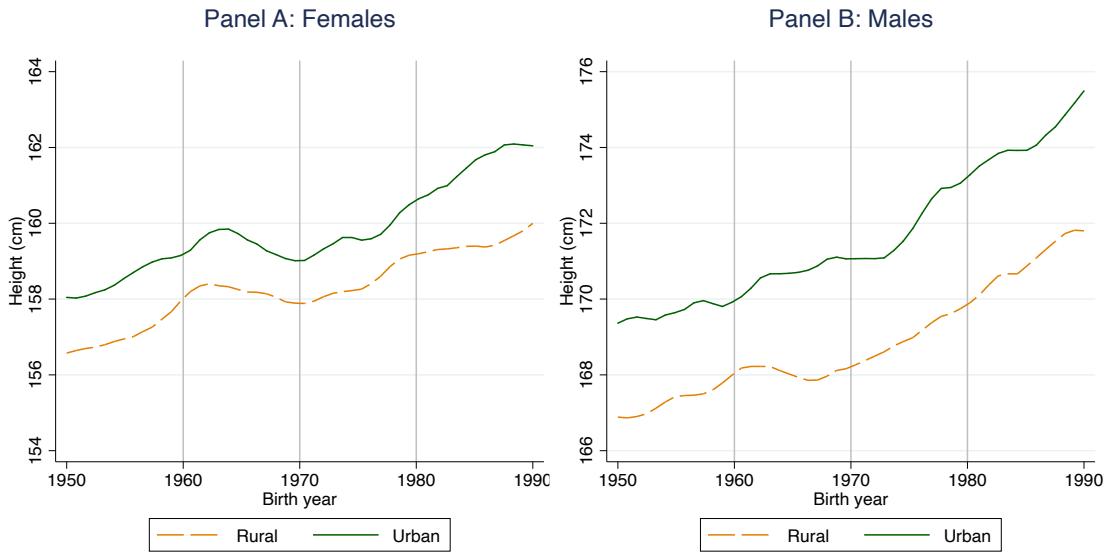
[6]It was not until the early 2000s that a New Cooperate Medical Scheme started to take shape.

period of study for this paper.

2.2.2 One Child Policy and son-preference

After the early success in implementing family planning policies (the 'Later, Longer, and Fewer' policies) to bring down fertility, a much more strict, 'One Child Policy' (OCP) was put in place in June 1979 [Center for Population Studies, 1986], although it only applied to urban populations. In rural areas, a second birth was allowed if the first child was a girl [Peng, 1991]. Out-of-quota births were subject to fines, but to different degrees depending on local officials' policy enforcement means and political lineage [Edlund et al., 2007]. The reason a relaxed policy was implemented in rural areas is closely linked to the fact that the son-preference culture is more prevalent in rural areas than urban China. In any case, the fertility restrictions were real and even with a much looser policy, issues such as abnormal sex ratio at birth became a very serious problem in the rural areas.

Figure 3-2: Height by rural-urban and gender: 1950-1990



Source: China Family Panel Survey; China Health and Nutritional Survey; Rural-Urban Migration in China
Notes: The graph is generated with local polynomial smoothing line with bandwidth of 1.2.

In general, when fertility drops, ceteris paribus, nutrition level increases, and height increases. However, under a strong culture of son-preference and a potential financial penalty for having an additional birth, the change in height for boys and girls may differ. Indeed, this may well be what happened in rural China. Figure 3-2 shows the average height by rural/urban status and by gender.[7] The figure indicates that rural men and women on average are shorter than their urban counterparts who were born in the same year, and more so for

[7]Due to small sample size, especially for the urban sample, we merge three data sets to plot Figure 3-2: the China Family Panel Survey, the China Health and Nutritional Survey, and the Rural-Urban Migration in China.

men (around 4 cm difference between urban and rural men) than for women (around 2 cm difference between urban and rural women). However, if we consider the long-term trend, rural men track urban men quite well over the entire period, but rural women's growth in height stalled after the introduction of the OCP (cohorts born after the late 1970s).[8]

3 Model specifications and estimation issues

Although Figure 3-2 presents the unconditional time trend of how Chinese adult stature evolved between 1950 and 1990, the unconditional means may be confounded with compositional changes in population over time. Thus, to answer our first question, and to quantify the height changes, we estimate a simple regression, controlling for birth-county fixed effects. Specifically, we estimate two different specifications of the following equation:

$$H_{ic} = \alpha_0 + \alpha_1 T_i + \delta_c + \varepsilon_{ic} \tag{1}$$

where H_{ic} is the adult height of individual i born in county c and measured at the survey year. T_i in the first specification is a linear time trend, generated from individuals' birth years. In the second specification, T_i is measured as three decade dummy variables, indicating the decade the individual was born: the 1960s, the 1970s, and the 1980s, while the 1950s is the omitted category. δ_c is the birth-county fixed effects and ε_{ic} is the random error.

To understand how local economic and social environments affect individuals' stature, we estimate the following equation:

$$H_{ic} = \beta_0 + \beta_1 \sum_{n=1}^{3} \omega_n E_{cn} + \beta_2 X_{i14}^{P} + \beta_3 W_{ic} + \beta_4 Sib_{ic} + \lambda_c + \epsilon_{ic} \tag{2}$$

where H_{ic} is the adult height of individual i who were born in county c and was measured at the survey year. n takes a value between 1 and 3 which indicates the first three years of the child's life (aged between 0 and 3). E_{cn} is a vector of socioeconomic conditions in county c at time when the individual was aged n, while ω_n is a damping coefficient, which varies across different age. In this study ω_n is calculated based on the WHO height-for-age median value growth chart. We discuss the details of how ω_n is calculated in Appendix B. Thus, in this regression we replaced a birth year trend, T with a vector of local socioeconomic conditions which also varies over birth year to see how the over-time variations of these conditions affected the average height of the population. E consists of four different measures for socioeconomic conditions at the county level: log of industrial and agriculture outputs per capita, separately, (we also use the variable 'total output per capita' to test the sensitivity), log of government social expenditure per capita, and mortality rate.[9] X_{i14}^{P} is a vector of parental

[8] The fast catching up of rural female height to that of the urban female during the famine years may be related to the survival bias discussed in Gørgens et al. [2012].

[9] Here we are generating weighted average local socioeconomic conditions during the person's height growing period. Ideally, we would want to take into account the entire height growth period (age 0-16) when estimate

characteristics at the time the individual was aged 14,[10] including parental education, whether father was a party member, and the rank of father's occupation. This parental information during individuals' childhood can capture family income information over and above local socioeconomic condition effect. The vector W_{ic} consists of variables indicating individual i's ethnicity, gender, and birth order at the survey year, while Sib_{ic} is the number of siblings individual i has. λ_c is the birth-county fixed effects and ϵ_{ic} is a random residual.

In equation (2) we try to handle the potential endogeneity problem by using the lagged economic and social conditions together with the county fixed effects. We argue that controlling for time-invariant county factors, it is unlikely that individuals' adult height could affect the socioeconomic conditions in their childhood (reverse causality issue). There may be omitted variable problems. For example, parental height is an important determinant of child's height but we did not control for it as only a very small proportion of observations has this information. There is a possibility that parental height could also be related to socioeconomic conditions of the county at the time of the individual's childhood. To the extent that this slight possibility exists, $\widehat{\beta}_1$ in equation (2) may be estimated inconsistently. However, as the socioeconomic conditions are recorded at the time the individual was a child and at that time their parents' heights had long stopped growing, it is less likely that parental heights and the local socioeconomic conditions measured at that time are correlated strongly.

In this situation, parental heights can only be correlated with the 'socioeconomic condition' variables if there were strong persistence in these variables. We have two reasons to suggest that this is unlikely to affect our estimation. First, considering China's economic growth path, as discussed in Section 2, it is clear that local socioeconomic conditions have changed dramatically. Second, to the extent that there is a certain degree of persistence, the county-level fixed effects should have taken care of it. Nevertheless, we test the effect of this omission in the sensitivity test section (Section 6).

Equation (2) assumes that the impact of socioeconomic conditions is linear: the effects are constant over the forty-years period. To test if this is the case, and more importantly, to identify the economic reform effect, we introduce an exposure measure, which indicates during individuals' stature growth period (age 0-16) what proportion of that period was during the economic reform period (1979 and afterward). Specifically, those who were born from 1979 onwards would have spent 100% of the stature growth period under the introduction of economic reform, and those born between 1964 and 1978 would have spent a various proportion of their stature growth period under the economic reform period, while those born before

equation (2). However, two issues stopped us from doing so. First, to take 16 years of weighted average of these variables we need to have data extending to 2006 for those who were born in 1990, but our data end at 1990. If we do take 16 years of weighted average, we would lose a large proportion of observations from our sample. Second, taking 16 years of weighted average will significantly reduce the variation of these variables cross different cohorts and hence reducing the precision of our estimates. Due to these reasons we decide to take the weighted averages of the first three years, which are regarded as the most crucial years of height growth period and account for over 40% of the total height growth.

[10]It would be ideal if we have parental information at a much earlier stage of the child development as the stature development mainly occurs in the first three years of individual's life according to WHO height-for-age median value chart. However, the earliest information we can obtain from the survey is at age 14.

1963 were not exposed to these policy changes.[11] Including the exposure variable into equation (2) allows us to capture the net 'post-1979' effect. In this situation, the impacts estimated of socioeconomic conditions and other personal variables would capture the pre-1979 effects.

However, there is a complication. The economic reform was introduced in 1979, so as the OCP.[12] Thus, the exposure variable can only measure the net effect of the combination of the two policies. Our purpose is to identify the economic reform effect and to answer our third question of how the OCP and accompanying economic sanctions towards violation of the policy may have affected the two gender groups differentially due to the son-preference culture. To do so we need to separately identify them. Our strategy is to interact the exposure variable with the number-of-sibling variable. Because we control for both the level effects of 'exposure' and 'number-of-siblings' variables, the interaction term captures the sibling effect during the post-1979 OCP policy, while the exposure variable captures the economic reform effect. We estimate this specification separately for male and female samples to gauge the differential effect of OCP on the two gender groups.

4 Data and descriptive statistics

4.1 Data

We use two main data sources: the China Family Panel Studies (CFPS) initial 2010 wave complemented with its 2012 and 2014 waves, and the county-level gazetteer data we hand collected from hundreds of books. The CFPS is an ongoing panel survey conducted by Peking University.

The CFPS 2010 wave surveyed 33,600 respondents. The survey covers a rich array of individual and household level information, including individual self-assessed anthropometric measurements. Our study investigate height of individuals born between 1950 and 1990 and this restriction reduces our sample to a total of 25,243 individuals. Further excluding individuals with missing values relevant to our estimation, we have a final sample of 18,508 individuals. While our sampling frame is based on the 2010 wave, to reduce reporting errors on height, we take the mean height reported in the three survey years (2010, 2012 and 2014). Further, to enlarge our sample we include individuals who reported height at least once in the three waves. We also exclude obvious outliers from our final sample.[13]

The height variable in the CFPS survey is self-reported. Comparing self-reported height

[11] We discuss the details of this exposure measure in the next section.

[12] Note that both economic reform and OCP have different channels through which they can affect height of individuals. Economic reform increased household income, but in its early years it reduced rural household access to health care. The situation did not improve much until the late 1990s. The net effect of the reform should be a combination of these two potentially opposite effects. Similarly, the introduction of the OCP should reduce sibling numbers, hence increase resources to be shared among siblings within the family. But for households that prefer sons over daughters, the out-of-quota birth in order to get a male birth would increase sibling number. What is more, if an out-of-quota birth occurs it would incur a financial fine for the family and further reduces the resources to be shared among siblings.

[13] Appendix Table 3-A1 reports details as to how many observations were excluded based on each of the sample restrictions.

by birth cohort from the CFPS with the measured heights for the same cohorts from the China Health and Nutrition Survey (CHNS) we find that the differences between self-assessed and measured heights are not large.[14] However, the height-by-age measure could be subject to other biases. One is that there could be positive selection among those that survive to older ages, which could create upward bias among older cohorts. In the case of China, we need to worry about this for those born just before and during the famine years.[15] For cohorts born after the early 1960s it should not be a problem. Another measurement-related issues is that height decreases in older age, largely due to compression of the spinal column, which would lead to an downward bias in earlier cohorts.[16] Both of these potential measurement issues point to our 1950s cohorts as being problematic, but the direction of the bias from the two issues are opposite and could to some extent cancel out each other. Further, Fernihough and McGovern [2015] has indicated that the potential shrinkage in age 50 to 60 is quite small and stable. To allow our analysis for a longer period we opted to keep the 1950s cohorts in the sample, but test sensitivities of this inclusion to our main conclusions.[17]

An important feature of the CFPS is that it records each respondent's birthplace and household registration status in childhood at the county level. This means that we can fairly accurately capture their birth location and link individuals to local socioeconomic conditions. We divide our final sample into two subsamples: those whose household registration at age 3 were rural (16,051) and those who were registered as urban (2,457). Given the vastly different economic institutions for the rural and urban regions as discussed in the background section, our analyses are always conducted separately for rural and urban samples.

To link individuals to local socioeconomic conditions in early childhood, we merge the CFPS data with data collected from local gazetteers using birth year and birth county. Our sample individuals were born into 159 counties of 25 provinces. The data from county gazetteers include annual agricultural and industrial output values (and total output values), government social expenditure (mainly including expenditure on health and education), mortality rate, and total population over the years from 1950 to 1990. According to Almond et al. [2019], data from the gazetteers are similar to those from the yearbooks but are more accurate when they disagree. However, not all the counties have all the variables for every year between 1950 and 1990. We use yearbook data to fill in some of the missing values, but that still leaves a rather large proportion of missing values. To ensure that the sample size is not reduced too much, we use available years of county-level data together with the prefecture or provincial time trend to predict for the years where the county-level data are missing. More details of data source, collection, and prediction are available in Appendix C.

An additional issue related to merging of county-level gazetteers data with individual-level

[14] See Section 6.1 for discussion

[15] Gorgens et al. (2012) estimate that selection during the Great Chinese Famine increased height by about 1-2cm. On the other hand, the so-called the scarring effect decreased height by about the same amount [Gørgens et al., 2012].

[16] Huang et al. [2013] estimate the height shrinkage for Chinese men and women from the age of 60 onwards at about 2 cm per decade.

[17] See Appendix Table 3-A2.

data need to be discussed here. In China each region has two type of household registration (*hukou*): 'rural' or 'urban'. A majority of the population in rural areas have 'rural' household registrations, but a small proportion also has 'urban' household registration, such as local mayors or a small number of public servants. In urban regions, majority of population has 'urban' household registration, but there are also 'rural' *hukou* people living in the periphery of the city. However, with regard to local socioeconomic conditions, such as agriculture output, industrial output, etc., for the same county or prefecture there are only one set of aggregated information. Thus, we merge county-level variables with individuals' birth-county identification, regardless of their *hukou* status. That is to say that in the same county or city everybody is merged to the same local socioeconomic condition variables relevant to their birth year regardless of their *hukou* status. In the regression, however, we include a variable which captures each county's rural or urban *hukou* population share. The information on rural or urban *hukou* populations at the county level are also collected from the county gazetteers.

4.2 Descriptive Statistics

Table 3-1 reports the summary statistics for the rural and urban samples, separately. As height development mainly occurs in one's childhood, we define 'rural' and 'urban' sample based on individuals' *hukou* registration status at age 3.[18] Panel A of Table 3-1 reports individual and parental characteristics, and Panel B presents the socioeconomic conditions of the individuals' birth counties at their birth year.

Panel A of the table indicates that, on average, the height of our rural male sample is 169 cm, and that of the urban sample is 172 cm, a 3 cm difference. For females, the difference is around 2 cm (=161-159). However, as indicated in Figure 3-2, the differences vary by birth cohorts: it is larger for the younger birth cohorts than for the old birth cohorts. The average age of the rural and urban samples are about the same, with the rural sample being around 0.7 years older than that of the urban sample. The average number of siblings is 3 among rural individuals and 2 for urban individuals.

One key element of family background is parental education. The average years of schooling of the parents (the average of mother and father) is 2.9 years for rural individuals and 6.2 years for urban individuals. To better capture family income level when our sample individuals were young, we generate a variable of father's occupation rank when the child was 14 years of age. The occupation classification in the CFPS is very detailed. There are total of 364 occupations in the sample. Among agriculture workers the total number of occupation is 28. We rank these occupations based on the mean earnings in 2010 for each occupation.[19] The higher the rank of occupation, the higher the average earning. The average rank of father's occupation is 66 in the rural sample and 191 in the urban sample, which is about three times higher than the rural sample. About 11.4% and 21.1% of individuals' fathers were the

[18] Around 15% of our sample switched *hukou* status from rural to urban between age 3 and current.

[19] It is more accurate to use the average income level of each occupation at child's age 14, rather than as of 2010 but we do not have that information.

communist party members in the rural and urban samples, respectively. The individual's own education differs sharply between rural and urban individuals by about 4.5 years of schooling. Finally, the proportion of Han Chinese is over 90% in both samples.

Panel B of Table 3-1 shows that people in rural areas faced poorer socioeconomic conditions compared to those in urban areas. First, the total output per capita is lower. Per capita total industrial and agricultural outputs in the rural sample are 650 yuan in 1990 prices, which is far lower than that of the urban sample (3,690 yuan).[20] The local government social expenditure per capita is 10 yuan for rural people, and 20 yuan for urban people. The average mortality rate is higher among rural residents compared to urban people.

5 Results

5.1 Long-term trends in the heights of Chinese

This section explores the long term trend in adult heights of individuals born between 1950 and 1990. As shown in Figure 3-2, over the 40-year period, male height increased by around 6-7 cm, and female height increased by around 4 cm.[21] A common pattern across the four groups is that slopes are steeper among post-1970 birth cohorts relative to pre-1970.

However, the unconditional trend of the mean height presented in Figure 3-2 may confound with compositional changes in population over time. Thus, we estimate time trends in a simple regression, controlling for birth-county fixed effects as indicated in equation (1), and see the average growth within a county across cohorts. As discussed in Section 3 we estimate two different specifications of equation (1), where time is measured as: (1) a linear trend and (2) three decade dummy variables.

Table 3-2 presents the results from the first two specifications. Panels A and B present results for the rural and urban samples respectively. Columns 1 and 3 present results which treats over time change of height as a linear trend for males and females separately. The results in Panel A indicate that the annual height increase is 0.127 cm for rural males, and 0.070 cm for rural females. Panel B shows that the annual height growth is 0.153 cm for urban males and 0.104 for urban females, which are approximately 0.03 cm higher relative to the rural sample. Since birth-county fixed effects are included, this growth reflects within-county growth patterns. Thus, over 40-years and controlling for regional variations, rural and urban male height increased by 5.08 and 6.12 cm, respectively, while rural and urban females increased by 2.80 and 4.16 cm, respectively. Relative to the one cm height increase in a decade in the 20th century for Western European males [Steckel, 1995], the height growth of the Chinese male population between 1950 and 1990 seems to have been greater. However, here

[20] The reason why the urban sample also recorded agriculture output is that there are also rural areas within each prefecture. One possible reason that the value of agriculture output is higher for the urban sample than for the rural sample perhaps is.

[21] A cautionary note here: The height of those born in the 1950s were in their 50s in 2010. Studies on height shrinkage have suggested that human height shrinks from around age 40 and the process becomes more progressive after around age 50 [see, for example, [Fernihough and McGovern, 2015].

Table 3-1: Descriptive statistics

	Rural *hukou*			Urban *hukou*		
Mean (Std. Dev)	Total	Male	Female	Total	Male	Female
Panel A: Individual characteristics						
Height (cm)	163.85	168.79	158.83	166.69	172.20	160.79
	(7.37)	(5.73)	(5.13)	(7.82)	(5.63)	(5.05)
Birth year	1969	1969	1969	1970	1970	1969
	(11.05)	(11.04)	(11.05)	(11.81)	(11.74)	(11.89)
Age in 2010	41.16	41.31	41.00	40.48	40.27	40.70
	(11.05)	(11.04)	(11.06)	(11.80)	(11.73)	(11.89)
Male	0.50			0.52		
	(0.50)			(0.50)		
Number of siblings	3.00	2.92	3.07	2.01	1.96	2.06
	(1.85)	(1.86)	(1.84)	(1.86)	(1.83)	(1.89)
Birth order	2.56	2.56	2.57	2.15	2.16	2.14
	(1.59)	(1.59)	(1.60)	(1.48)	(1.47)	(1.49)
Parental average schooling years	2.94	2.95	2.92	6.20	6.24	6.17
	(3.23)	(3.23)	(3.23)	(4.29)	(4.32)	(4.25)
Father's occupation rank at child's age 14	66.14	65.40	66.89	191.23	190.06	192.49
	(80.27)	(79.49)	(81.04)	(79.52)	(79.40)	(79.66)
Father's party membership at child's age 14	0.11	0.11	0.11	0.21	0.21	0.22
	(0.32)	(0.32)	(0.32)	(0.41)	(0.40)	(0.41)
Individual schooling years	6.83	7.70	5.95	11.40	11.45	11.33
	(4.37)	(3.98)	(4.58)	(3.36)	(3.25)	(3.48)
Han Chinese	0.90	0.90	0.91	0.96	0.96	0.96
	(0.29)	(0.29)	(0.29)	(0.19)	(0.19)	(0.20)
Panel B: County characteristics[1)]						
Total output per capita[2)]	0.07			0.37		
	(0.10)			(0.39)		
Agricultural output per capita[2)]	0.03			0.03		
	(0.02)			(0.07)		
Industrial output per capita[2)]	0.03			0.33		
	(0.09)			(0.38)		
Gov. expenditure on social services per capita[2)]	0.001			0.002		
	(0.002)			(0.003)		
Death rate (‰)	9.52			7.27		
	(5.63)			(4.45)		
N	16051	8098	7953	2457	1270	1187

Note:
1) County-level variables are weighted by the rural *hukou* population share for the rural sample and urban hukou population share for the urban sample;
2) Unit: 10000 yuan; measured in 1990 price level;

we assumed a linear growth path of height and based on the discussion in the background section, economic and social environment over the 40-years period has changed dramatically. The height increases should also vary accordingly.

Table 3-2: Time trend in height growth

	Rural Total		Rural Male		Rural Female	
	(1)	(2)	(3)	(4)	(5)	(6)
Time trend	0.091***		0.127***		0.070***	
	(0.006)		(0.006)		(0.006)	
Born in 1960s		0.785***		0.944***		1.122***
		(0.138)		(0.153)		(0.157)
Born in 1970s		1.416***		1.937***		1.369***
		(0.177)		(0.199)		(0.167)
Born in 1980s		2.859***		3.947***		2.304***
		(0.202)		(0.203)		(0.188)
Constant	162.140***	162.672***	166.401***	167.218***	157.498***	157.657***
	(0.120)	(0.112)	(0.118)	(0.118)	(0.111)	(0.109)
Observations	16051	16051	8098	8098	7953	7953
Adjusted R^2	0.097	0.096	0.225	0.222	0.164	0.163

	Urban Total		Urban Male		Urban Female	
	(1)	(2)	(3)	(4)	(5)	(6)
Time trend	0.149***		0.154***		0.107***	
	(0.011)		(0.012)		(0.009)	
Born in 1960s		0.634		0.589		0.721**
		(0.382)		(0.422)		(0.359)
Born in 1970s		2.326***		1.856***		1.448***
		(0.497)		(0.526)		(0.540)
Born in 1980s		4.293***		4.378***		3.346***
		(0.350)		(0.408)		(0.361)
Constant	163.775***	164.892***	169.151***	170.463***	158.710***	159.422***
	(0.212)	(0.243)	(0.233)	(0.296)	(0.176)	(0.249)
Observations	2457	2457	1270	1270	1187	1187
Adjusted R^2	0.090	0.087	0.223	0.214	0.118	0.117

Note:
1) Robust Standard Errors (SEs) are presented in parentheses;
2) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
3) County and prefecture fixed effects are included for the rural and urban samples, respectively;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Columns 2 and 4 include three decade dummy variables, hoping to capture the effect of economic and social environmental changes over different decades on height. The omitted category in each case is those who were born in the fifties. Thus, as each of the 1960s to 1980s decade is compared to the same base, their relative magnitudes should not be affected by the shrinkage issue caused by the 1950s cohort being in their 50s. Panel A shows that rural males born in the 1960s are 0.944 cm taller than the reference category of the 1950s, while the difference for rural females is 1.122 cm. For the urban sample (Panel B) we find that the 1960s males and females are 0.589 and 0.721 cm taller than their 1950s counterparts, respectively, but the difference is statistically significant only for the female sample. The fact that female 1960s

groups are doing better than their male counterparts when compared to those born in the 1950s could be related to the shrinkage issue. Literature has found that the height shrinkage for women is worse than that for men (see, for example, Huang et al. [2013]; Fernihough and McGovern [2015]).

The 1970s groups are taller than the 1960s groups for both rural and urban samples. The difference for rural males is 0.993 cm (=1.937-0.944), for rural females it is 0.247 cm (1.369-1.122), for urban males it is 1.267 cm (1.856-0.589), and for urban females it is 0.727 cm (1.448-0.721). The differences between the 1970s and 1980s groups are 2.010 cm for rural males, 0.935 cm for rural females, 2.522 cm for urban males, and 1.898 cm for urban females. All of these differences are statistically significant with the exception of urban females between 1960s and 1970s.[22]

The results obtained from Table 3-2 suggest that, in general, the average height growth for Chinese population in the 40 years has improved a lot. Three out of four sub-samples (rural and urban males and urban females) have had above 1 cm per decade growth. Most of the growth occurred after the 1970s, and, in particular, the 1980s had the strongest growth for all four groups. This corresponds quite well with our expectations as the economic reforms started in late 1970s have substantially improved living standards in China. What is also evident from Table 3-2 is that the urban population, on average, is taller than the rural population (the constant terms in Table 3-2 indicate that urban males and females who were born in the 1950s are about 3 cm and 2 cm taller than their counterparts born in rural areas, respectively[23]) and also had more height growth than their rural counterparts. The final observation from Table 3-2 is that the average height of males had increased more than their female counterparts over time. More importantly, the gender gap in height is more profound in rural than urban areas and it is increasingly so. This finding is consistent with Schwekendiek and Baten [2019] who also show that a pro-male bias in height is larger among the 1980s birth cohort than the 1970s, especially in rural areas.

China's rural-urban divide policy can probably explain a large part of the difference in the level and increment of the average height between rural and urban populations. This is not only because urban average per capita income is almost three times of that of rural average per capita income and urban people consume more calories (see Figure 3-1), but also because healthcare provision has always been more widespread and better quality in urban than rural areas as discussed in Section 2.

The increasing difference in height between men and women over time, especially for the rural population, is likely to be related to the introduction of the OCP under the culture of son-preference. The late 1970s and the 1980s saw the introduction of OCP which reduced the

[22]The significance test results are available upon request from the authors.

[23]When comparing the constant terms across regressions one needs to be aware of the omitted categories of multiple category dummy variables are consistent across the regressions. Luckily, the only group of dummy variables we include in these regression (apart from dummies for decades) is the county-fixed effects. We have estimated these regressions without county-fixed effects and the results show that they do not affect the constant terms much. These results are available upon request from the authors.

fertility rate. In the same period, the introduction of the Household Responsibility System in rural areas significantly increased household incomes. Both of these events should be positively related to the average height as they increased available resources to each child in the household. However, the benefits seem to have been reaped by males but not females. This is likely to be related to the traditional son-preference under the OCP. Under the son-preference culture, traditional families (more likely to be rural households) are more likely to provide more to boys than to girls. Thus, within a family with mixed-gender siblings girls are likely to be discriminated against in resource allocation. In addition, in rural areas, the OCP allowed families with firstborn girls to have a second child. As a result, females are more likely to be in households with more children. If Becker's Quantity-Quality trade-off effects are at work, ceteris paribus, we would also expect females be further discriminated in resource allocation [Becker and Lewis, 1973]. The combined effect would see female height increases slower than male height despite the reduction in fertility rate and the increase in per capita income during the same period.

5.2 Height and its contributing factors

5.2.1 Socioeconomic factors

Although columns 2 and 4 of socioeconomic Table 3-2 have shown that the period where average height growth is the most is in the late 1970s and the 1980s, what exactly contributed to such a growth is unknown. In this subsection we turn to our second question to examine how local living condition changes are associated with the height-growth pattern. In doing so, we estimate equation (2). As previously discussed, equation (2) examines the effect of one's birth county's socioeconomic conditions in the first three years of one's life on the height of individuals. To better capture the birth-period effect, we examined WHO height-for-age median value for boys and girls separately and discovered that one's height grows the most in the first three years of life, accounting for almost 40% of the total height.[24] We thus use the share of the first three years' height growth rate as the weight to adjust the social economic conditions at the time of individual's first three years to generate weighted average of the three years conditions.[25]

Table 3-3 presents the estimation results for the rural total, male and female samples (columns 1 and 3) and the urban total, male and female samples (columns 4 and 6). Controlling for parental education, occupation, and party membership status when the person was a child (aged 14 years), individual and family characteristics, together with the birth-county fixed effects, we find that local conditions at one's first three years of life play important roles in individuals' adult height. Every 10% increase in agriculture and industrial outputs increase average height by 0.043 cm and 0.021 cm, respectively, and the coefficients are statistically significant at the 5% and 1% levels, respectively. These are rather large impacts. During this

[24] See Section 3 for more discussions regarding why we focus on the first three years.
[25] The details of how the weight is calculated are presented in Appendix B.

period, China's average agriculture and industrial output increased by 3.1% and 8.0% per annum, respectively [NBS, 1999]. In particular, during the economic reform period of the 1980s, the annual growth rates were 6.0% and 7.6%. These large effects, however, seem to have mainly affected male average height, but not female. A 10% increase in agriculture and industrial outputs increase average male height by 0.075 cm and 0.026 cm and both are statistically significant at the 5% level, but the effects on average female height are almost zero and none is statistically significant (the joint significance of the two variables is also not statistically significant with a p-value equal to 0.26). Table 3-A3 in Appendix A includes total output per capita instead of the two separate output variables and shows the consistent pattern.

Government social expenditure per capita (mainly on education and health) at the time of individuals' first three years of life is an important contributor to both rural male and female average heights. However, the effect is almost twice the size for males than for females. An additional 10% increase in social spending increases male average height by 0.065 cm and female average height by 0.033 cm. The general health condition at the time of individuals first three years of life measured by county-level mortality rate, though, is only a statistically significant contributor for rural females. A 10% reduction in mortality rate is associated with a 0.0053 cm increase in rural female average height. These results that local socioeconomic environment contributes more to rural male average height change than to rural female average height change. Indeed, including the four county-level variables in rural male and female regressions explained the average height of rural males and females by 23% and 16%, respectively (the adjusted R^2s are 0.23 and 0.16). If we exclude the four social/economic environment variables, the adjust R^2 for the rural male sample reduces by 2.5 percentage points to 0.20 (or 11% of the total explained variation in individuals height can be attributed to these social/economic factors), whereas for the rural female sample, it only drops by 1 percentage point to 0.155 (6.1% of the total explained variations).

For the urban samples, output variables have limited effects on average height, except for a small positive effect of the industrial output on average height of urban females. The effect of government social spending per capita, however, seems to be very important on height of urban people. A 10% increase in government social spending increases average urban height by 0.089 cm. The effect is particularly strong for males, a 10% increase in social spending increases average urban male height by 0.113 cm, whereas for females, it increases the average height by 0.054 cm. The level of government social spending per capita for our sample prefectures increased by 8.4% per annum between 1950 and 1990. For the last decade, the annual increase was much higher at 13.9% per year. Reduction in mortality rate plays no role in average urban population height, either for males or females. Once again, the four regional socioeconomic condition variables, in particular the government social spending, contributed a rather large portion of explained height variation for the male sample (11.8%=0.0257/0.218) but was small for the female sample (5.3%=0.006/0.119).

Parental characteristics at the time the person was aged 14 are used to capture family socioeconomic background when the individual was young. We have three variables: the

Table 3-3: Effects of regional social and economic factors on height

	Rural sample			Urban sample		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Agricultural output p.c.- age 0 to 3	0.427**	0.746***	0.074	0.039	0.212*	-0.155
	(0.168)	(0.233)	(0.217)	(0.106)	(0.125)	(0.137)
Log Industrial output p.c.- age 0 to 3	0.206***	0.259**	0.158	0.066	-0.057	0.177*
	(0.077)	(0.101)	(0.103)	(0.089)	(0.167)	(0.106)
Log Gov. Social expenditure p.c.- age 0 to 3	0.477***	0.646***	0.332***	0.893***	1.134***	0.542***
	(0.099)	(0.136)	(0.115)	(0.092)	(0.123)	(0.147)
Death rate (‰) - age 0 to 3	-0.032**	-0.015	-0.053***	0.019	0.018	0.014
	(0.012)	(0.015)	(0.016)	(0.037)	(0.047)	(0.052)
Parental schooling years	0.077***	0.102***	0.049**	0.127***	0.169***	0.123***
	(0.016)	(0.023)	(0.022)	(0.033)	(0.039)	(0.042)
Father's occupation rank at child's age 14	0.002***	0.002***	0.002**	0.000	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Father's party membership at child's age 14	0.360***	0.299	0.382**	-0.093	-0.329	0.148
	(0.129)	(0.195)	(0.190)	(0.294)	(0.409)	(0.353)
Male	10.022***			11.340***		
	(0.096)			(0.245)		
Han Chinese	0.108	0.259	-0.005	-0.459	0.145	-1.149
	(0.250)	(0.365)	(0.329)	(0.848)	(0.842)	(1.020)
Number of siblings	-0.085**	-0.071	-0.055	-0.197*	-0.105	-0.294*
	(0.039)	(0.047)	(0.053)	(0.108)	(0.177)	(0.151)
Birth order	0.061	0.034	0.053	-0.058	-0.064	-0.042
	(0.037)	(0.050)	(0.050)	(0.121)	(0.219)	(0.177)
Observations	16051	8098	7953	2457	1270	1187
Adjusted R^2	0.560	0.225	0.164	0.610	0.218	0.119

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples;

3) Unit: log of 10000 yuan; measured in 1990 price level;

4) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values; for father's party membership and occupation are included;

5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

average years of schooling of mother and father (labelled as parental years of schooling); father's occupation rank when the child was 14 years of age based on the mean earnings of all the occupation in 2010; and whether the father was a communist party member when the child was 14 years of age. All three variables are statistically significant and positively related to individuals' height for the rural sample. For the urban sample, the only statistically significant parental variable is parental years of schooling.

Individual and other family characteristics are not associated with average height apart from the number of siblings. For both rural and urban full samples, sibship size is associated with individual's height negatively, which is consistent with the Quantity-Quality trade-off theory, as well as empirical literature on many other countries [Hatton and Martin, 2010; Liu, 2014; Hatton, 2014, 2015; Bailey et al., 2016; Kubo and Chaudhuri, 2017]. However, when we separately estimate for the male and female samples, none of the coefficients is statistically significant. As discussed previously, while the fertility policy in China has changed significantly in Table 3-3, we do not take the policy change into account.

5.2.2 Economic reform, One Child Policy and height

The above analysis assumes that over the past 40 years, factors affecting height have had the same effects. However, the regression results reported in columns 2 and 4 of Table 3-2 have clearly suggested that the height growth over the 40-year period was not linear and that the major growth occurred in the 1980s. The Background Section discussed how much changes took place at the end of 1970s: the introduction of the Household Responsibility System and the subsequent switching from a planned to a market-oriented economic system has led China to experience an unprecedented and sustained economic growth path; the abolition of the commune system has reduced social services to be provided in rural areas for a long time; and the introduction of the OCP which has substantially reduced fertility. Both of these important socioeconomic changes would have had important implications on height changes. An increase in income provides families with more resources for everyone, while at the same time the reduction in social services at the village level may play an opposite role. Similarly, the introduction of the OCP reduced fertility and enabled each child to get more resources, *ceteris paribus*. However, under the culture of son-preference and with the financial punishment of a violation of the policy being very high,[26] the net impact of the OCP is complicated depending on whether the individual is a male or female, and whether the family has a son preference.

The lack of social-welfare provision generated by the abolishment of the old commune system during the initial economic reform period mostly affected rural people. The potential negative impact of OCP was also mainly borne by rural people, in particular, rural girls. In general rural Chinese are more traditional and have stronger son-preference than urban dwellers. When one has a strong son preference, and when the firstborn is a girl, it is more

[26] According to Scharping [2003]; Ebenstein [2010], the fine for violation of OCP in the 1980s was between 20-360% of urban household annual income, which is normally 2 to 3 times of rural household annual income.

likely that parents would have a second birth and if the second is still a girl, they may continue to have a third and more births until they have a boy, whereas if the first or the second birth is a boy, parents are likely to stop. Thus, girls are more likely to have more siblings than boys under OCP. In addition, girls are more likely to be born in families that violate the birth quota and hence trigger a substantial fine, which in turn generates more financial difficulties for the family. Therefore, the introduction of the economic reform and the OCP might have contributed to the less height growth for the rural people, girls in particular, than their urban counterparts.

In this sub-section, we examine if the economic growth and the introduction of the OCP have significantly changed the height growth pattern. To do so, we generate a new variable, 'exposure', which intends to capture the share of an individual's height growing years (between ages 0 and 16) was at and after 1979, the year of the start of the economic reform and the introduction of the OCP. That is, the exposure variable measures the relative degree to which one was exposed to the reforms and OCP between ages 0 and 16. For example, this share is equal to 100% for individuals born in and after 1979. Individuals who were born between 1964 and 1978 will have a varying share of their first 16 years of life spent during the reform and OCP era. Those who were born before 1964 have the value of this variable set to zero. We calculate the number of years individuals spent in the reform/OCP era and adjust it by the WHO height-for-age growth chart. The details of the method for generating this variable is presented in Appendix B.

We first estimate equation (2) with this additional variable to examine the net impact of economic reform and introduction of OCP effect. We then add the interaction term of 'exposure' and the number-of-sibling variable to try to disentangle the OCP effect.

The selected results for the rural and urban samples are reported in Tables 3-4 and 3-5, respectively. The first three columns of each table report the original and the two new specifications for the total sample, while columns 4 to 6, and 7 to 9 present results for male and female samples separately. The results from both the rural and urban samples show that people who spent some of their first 16 years of life after the introduction of economic reform and OCP are on average taller than those who did not (compare columns 2, 5 and 8 with 1, 4 and 7). For the rural male sample, this difference is around 1.8 cm. That is to say, those who spent 100% of their first 16 years of life during the reform and OCP era are 1.8 cm taller than those whose entire height growing period was during the pre-reform and pre-OCP era. Those who spent a lesser share of their first 16 years during this period had less gain in their height. This difference for the rural female sample is smaller, but is still statistically significant, at around 0.6 cm. The difference for the urban male sample is the largest, at 2.5 cm. It is more than twice as large (1.4cm) for the urban female sample as than for the rural female sample.

What is interesting is that once we added the variable indicating one's exposure to the economic reform and OCP period during the major height growing phase of one's life, all of the previous statistically significant production output variables become statistically insignificant and the size of the coefficients reduce dramatically. The magnitude of the coefficients on

'per capita government social spending' are also reduced by around one third of the size relative to the estimation without controlling for the exposure variable. In addition, the effect of mortality rate on height for the rural female sample is reduced as well. These results suggest that perhaps the gains in population height due to local socioeconomic conditions are more important during the period after the economic reform and the introduction of OCP. Indeed, when we interact the exposure variable with the local social-economic condition variables, we observe, for the rural sample, that the slope for industrial output is steeper for the group whose first 16 years of life was under the reform era than for those who were born earlier (See Table 3-A4). For the earlier cohort, this coefficient is positive but not statistically significant. The effect of agriculture output is also positive for the younger group and fairly large, but not precisely estimated. In addition, the effect of mortality is larger for the younger than for the older cohorts and precisely estimated. The only variable which had similar effect is the per capita government spending on health and education, which have a similar effect on height before and after the economic reform. For the urban sample, however, the only statistically significant difference is the government per capita health and education spending. For both groups, the coefficient is positive, but it is statistically significant only for the after-reform group. It is also very interesting that when we interact these socioeconomic variables with the exposure variable, it enlarges the size of the exposure variable significantly. Perhaps, there are economic reform effects that is over and above its effect through an increase in output and reduction in mortality on height. Such additional effects could include the fact that the reforms increased the ability of households to allocate resources in their own interest.

If we regard the coefficient on the exposure variable in columns 2, 5 and 8 of Tables 3-4 and 3-5 as the net gain from the economic reform and the introduction of OCP and compare these gains to their respective sample mean height, we find the gain accounts for 1.1%(=1.80/169) of the mean height for rural males, 0.4%(0.61/159) for rural females, 1.5% for urban males, and 1.3% for urban females. It is clear that men gained more than women from these institution and policy changes, indicating a certain level of son-preference in society. Also, the gain is larger for the urban people than for the rural people suggesting the prevalence of rural-urban divide policy in favour of urban people. The level of son-preference seems to be higher in rural than in urban societies, which correspond well with our understanding of social norm differences between rural and urban regions in China.

5.3 The fertility effect

Another question we raised in this paper is how much contribution the fertility change has played in the significant increase in average height of individuals. As economic reform and OCP was introduced at around the similar time, it is challenging to disentangle the fertility effects from the reform effects. Studies have shown that fertility decline improves child health through the Quantity-Quality trade-off mechanism [Becker and Lewis, 1973; Liu, 2014; Kubo and Chaudhuri, 2017]. As family size decreases, the amount of food or other resources avail-

able for each child increase. This may lead to an increase in height. At the same time, the introduction of the OCP, especially in the 1980s, is accompanied by a significant financial fine to the households which violate their birth quotas.

Although the coefficients on the exposure variable in columns 2, 5 and 8 of Tables 3-4 and 3-5 indicate a ***net*** effect of the economic reform and the introduction of the OCP, it is important for our purpose to gauge the net effects of economic reform vs. the introduction of OCP. Thus, in columns 3, 6 and 9 we further control for an interaction term between the exposure variable and the number-of-siblings variable. We argue that the effect of OCP on height mainly go through fertility. Conditional on household income (which is captured mainly by variables of parental characteristics at the time the individual was 14 years of age), the fertility has implications on resources available to each sibling not only through Quantity-Quality trade-off of reduced number of siblings, but also through potential fines incurred for violation of the OCP birth quotas (the additional income effect). Thus it is likely that before and after the introduction of OCP the effect of fertility on height differs. The interaction term here captures this differential effect.[27]

The results in Table 3-4 for both rural male and female samples show that the interaction term between the exposure variable and 'the number of siblings' are negative and statistically significant at the 1% level. The coefficients are quite large. For the male sample, while the level term of number of siblings is very small, the interaction term suggest that an additional sibling one has after the introduction of the OCP reduces individual height by 0.452 cm. This seems to suggest that fines from the violation of the policy is indeed important component of the net OCP effect. The effect for the female sample is similar: no effect for the pre-OCP cohort, but a 0.304 cm reduction with an additional sibling after the introduction of the OCP. Note that even though the coefficient for the female sample is slightly smaller than that for the male sample, the average decade growth for rural males was 1.3 cm while for rural females it was 0.7 cm (linear time trend estimation from Table 3-2). Relative to these lower bases, the share of the sibling effect for females is actually a much higher proportion. Thus, relative to the decade growth of 1.3 cm for the rural male sample, the effect of the additional sibling accounts for 35% (=0.452/1.3) of its decade growth in height; while for the rural female sample it accounts for 43% (=0.304/0.7) of its decade growth of 0.7cm.

What is more, controlling for the potential OCP effect, the coefficients on the original exposure variable, now mainly captures the economic reform effect, become larger. Rural males exposed to the economic reform since birth are 2.75 cm taller than those who did not fully exposed during the first 16 years of life (pre-reform birth cohorts). The gap increases to 2.75 cm from 1.8 cm when not controlling for the OCP effect. For rural female the effect is

[27] Although the exposure variable includes the early birth cohort between 1964 and 1978, the level of share of these early births is small. In addition, as the effect of the introduction of the OCP on number of siblings should not only be felt not only by the birth cohorts born after the introduction but also their siblings born before that. We feel that the exposure variable interact with sibling numbers can roughly capture the sibling factor since the introduction of the OCP. We also estimated another specification using dummy variable for birth year at and after 1979 to interact with the number of sibling variable and the results are largely consistent with the results presented here. These results are available upon request from the authors.

doubled, giving an increase from 0.61 cm to 1.28 cm.

For urban samples, however, the effect of OCP does not seem to be considerable. Although the size of the interaction term for both urban male and female samples are not small, none of them is precisely estimated and the joint significance tests between the 'sibling' level effect and the interaction term for both samples are not significant either. Controlling for the OCP effect, the coefficient on exposure increased slightly from 2.5 to 2.7cm for males and 2.1 to 2.3cm for females.

In Table 3-2, using non-linear time trend, we find a 2 cm height growth for rural males and 2.5 cm for urban males, an 0.9 cm for rural females and 1.9 cm for urban females. Now that we have controlled for the OCP effect, the true net economic reform effects for all groups are much higher (2.7 cm for both rural and urban males, 1.3 cm for rural females, and 2.3 cm for urban females). Thus, we may quantify that the OCP impact on height in general is negative: 0.7 cm for rural males, 0.4 cm for rural females, 0.2 cm for urban males and 0.4 cm for urban females. In sum, rural population suffered the most from the introduction of the OCP with regard to their long-term health measure — height.

Our results from columns 3, 6 and 9 of Tables 3-4 and 3-5 suggest that the net economic reform effect on average is as large in the rural areas as it is in the urban areas. But the OCP effect is much larger in rural areas than in urban areas. Further, the negative OCP effect is more important for rural females than that for rural males.

Table 3-4: Reform and OCP effects — Rural sample

	Rural Total			Rural Male			Rural Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Agricultural output p.c.- age 0 to 3	0.427**	0.130	0.106	0.746***	0.276	0.245	0.074	-0.081	-0.091
	(0.168)	(0.173)	(0.170)	(0.233)	(0.250)	(0.245)	(0.217)	(0.242)	(0.243)
Log Industrial output p.c.- age 0 to 3	0.206***	0.106	0.117	0.259**	0.114	0.120	0.158	0.102	0.113
	(0.077)	(0.084)	(0.083)	(0.101)	(0.104)	(0.102)	(0.103)	(0.108)	(0.108)
Log Gov. social expenditure p.c.- age 0 to 3	0.477***	0.345***	0.309***	0.646***	0.438***	0.402***	0.332***	0.262**	0.238*
	(0.099)	(0.098)	(0.099)	(0.136)	(0.138)	(0.139)	(0.115)	(0.122)	(0.123)
Death rate (‰) - age 0 to 3	-0.032**	-0.017	-0.027**	-0.015	0.010	-0.001	-0.053***	-0.045***	-0.052***
	(0.012)	(0.013)	(0.013)	(0.015)	(0.015)	(0.015)	(0.016)	(0.017)	(0.017)
Parental schooling years	0.077***	0.065***	0.060***	0.102***	0.086***	0.081***	0.049**	0.042*	0.039*
	(0.016)	(0.016)	(0.016)	(0.023)	(0.024)	(0.024)	(0.022)	(0.022)	(0.022)
Father's occupation rank at child's age 14	0.002***	0.002***	0.002***	0.002***	0.002***	0.002**	0.002**	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Father's party membership at child's age 14	0.360***	0.386***	0.413***	0.299	0.329*	0.368*	0.382**	0.400**	0.411**
	(0.129)	(0.130)	(0.131)	(0.195)	(0.195)	(0.196)	(0.190)	(0.191)	(0.191)
Han Chinese	0.108	0.120	0.099	0.259	0.248	0.250	-0.005	0.012	-0.020
	(0.250)	(0.254)	(0.259)	(0.365)	(0.367)	(0.362)	(0.329)	(0.332)	(0.338)
Number of siblings	-0.085**	-0.055	0.026	-0.071	-0.020	0.061	-0.055	-0.040	0.013
	(0.039)	(0.039)	(0.042)	(0.047)	(0.050)	(0.052)	(0.053)	(0.052)	(0.057)
Birth order	0.061	0.053	0.087**	0.034	0.016	0.060	0.053	0.051	0.069
	(0.037)	(0.037)	(0.037)	(0.050)	(0.050)	(0.051)	(0.050)	(0.050)	(0.050)
Male	10.022***	10.000***	9.987***						
	(0.096)	(0.096)	(0.094)						
Exposure		1.150***	2.118***		1.804***	2.748***		0.609*	1.280***
		(0.240)	(0.289)		(0.370)	(0.438)		(0.348)	(0.429)
Exposure × Number of siblings			-0.452***			-0.462***			-0.304***
			(0.073)			(0.116)			(0.095)
Observations	16051	16051	16051	8098	8098	8098	7953	7953	7953
Adjusted R^2	0.560	0.561	0.562	0.225	0.228	0.229	0.164	0.164	0.165

Note:
1) County fixed effects, a share of rural population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;
2) Robust Std. Err. are clustered at the county level;
3) * (**,***) denotes statistical significance at the 10% (5%, 1%) level;

Table 3-5: Reform and OCP effects — Urban sample

	Urban Total			Urban Male			Urban Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Agricultural output p.c.- age 0 to 3	0.039	-0.160**	-0.160**	0.212*	-0.001	-0.002	-0.155	-0.341**	-0.344**
	(0.106)	(0.071)	(0.069)	(0.125)	(0.090)	(0.089)	(0.137)	(0.132)	(0.131)
Log Industrial output p.c.- age 0 to 3	0.066	0.006	0.005	-0.057	-0.111	-0.115	0.177*	0.115	0.117
	(0.089)	(0.112)	(0.110)	(0.167)	(0.195)	(0.190)	(0.106)	(0.095)	(0.094)
Log Gov. social expenditure p.c.- age 0 to 3	0.893***	0.418***	0.401***	1.134***	0.629***	0.612***	0.542***	0.096	0.073
	(0.092)	(0.121)	(0.124)	(0.123)	(0.193)	(0.197)	(0.147)	(0.235)	(0.230)
Death rate (‰) - age 0 to 3	0.019	0.042	0.034	0.018	0.044	0.036	0.014	0.033	0.025
	(0.037)	(0.032)	(0.032)	(0.047)	(0.044)	(0.045)	(0.052)	(0.050)	(0.048)
Parental schooling years	0.127***	0.095***	0.095***	0.169***	0.136***	0.135***	0.123***	0.091**	0.092**
	(0.033)	(0.031)	(0.031)	(0.039)	(0.038)	(0.038)	(0.042)	(0.040)	(0.040)
Father's occupation rank at child's age 14	0.000	0.001	0.001	-0.001	-0.000	-0.000	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Father's party membership at child's age 14	-0.093	0.028	0.022	-0.329	-0.244	-0.265	0.148	0.315	0.319
	(0.294)	(0.299)	(0.299)	(0.409)	(0.405)	(0.402)	(0.353)	(0.375)	(0.378)
Han Chinese	-0.459	-0.357	-0.370	0.145	0.164	0.102	-1.149	-0.985	-0.961
	(0.848)	(0.824)	(0.819)	(0.842)	(0.799)	(0.800)	(1.020)	(1.015)	(1.022)
Number of siblings	-0.197*	-0.090	-0.065	-0.105	0.027	0.052	-0.294*	-0.209	-0.183
	(0.108)	(0.104)	(0.108)	(0.177)	(0.180)	(0.188)	(0.151)	(0.148)	(0.150)
Birth order	-0.058	-0.075	-0.054	-0.064	-0.087	-0.052	-0.042	-0.054	-0.040
	(0.121)	(0.119)	(0.116)	(0.219)	(0.217)	(0.201)	(0.177)	(0.173)	(0.174)
Male	11.340***	11.260***	11.261***						
	(0.245)	(0.237)	(0.238)						
Exposure		2.265***	2.462***		2.488***	2.715***		2.070***	2.266***
		(0.368)	(0.392)		(0.637)	(0.692)		(0.770)	(0.765)
Exposure × Number of siblings			-0.274			-0.331			-0.265
			(0.204)			(0.252)			(0.327)
Observations	2457	2457	2457	1270	1270	1270	1187	1187	1187
Adjusted R^2	0.610	0.613	0.613	0.218	0.225	0.225	0.119	0.125	0.124

Note:
1) Prefecture fixed effects, a share of rural population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;
2) Robust Std. Err. are clustered at the county level;
3) * (**,***) denotes statistical significance at the 10% (5%, 1%) level;

6 Robustness check

6.1 Self-reported height

Our dependent variable, adult height, is a self-reported variable. Existing evidence suggests that self-reported height could be systematically biased [P. et al., 1981; Rowland, 1990; Lu et al., 2016]. Supposing that an over-reporting problem is more prevalent among younger generations and high income areas, our estimates of economic growth effects on height could be driven by that.

To reduce such a concern, in our main estimations we have used individual average heights reported in different years - 2010, 2012 and 2014 - as the main dependent variable, hoping to wash out some of the reporting errors. In this subsection we further test the potential problem generated due to reporting errors on height by using a survey which directly measures height of individuals by interviewers — the China Health and Nutrition Survey (CHNS). As with the CFPS, the CHNS is an ongoing panel survey, although it commenced much earlier, in 1989. Unlike the CFPS survey, which covers 25 provinces, the CHNS only surveys households from 9 provinces. Also, the CHNS does not provide the household registration status and many other individual and parental information at one's birth or in childhood. These are the main reasons we do not use the CHNS in our main estimation. Further as we do not have individual household registration status (*hukou*) at birth, our rural and urban samples cannot be defined in the same way as what we did using the CFPS. We instead use the current household registration status information. Appendix D describes how the CHNS data are compiled.

Panels A and B of Table 3-6 compare the estimated results for equation (1) using the CFPS and CHNS data for rural male and female samples, respectively. We only include individuals born in the 9 provinces that the CHNS surveys. The results indicate that there is not much of differences between the estimations using the CFPS and CHNS.

6.2 Endogeneity of fertility

Up until now we have assumed the 'number of siblings' is exogenous. However, it is very likely that the number-of-siblings is endogenous. It is very likely that some unobservable parental characteristics determined both the individual's height and the number of siblings he/she has. For example, parental son-preference may affect how many children parents choose to have as well as how much resources their male/female children are allocated. If so, our estimates using the OLS method would generate inconsistent results on the number-of-sibling variable. To test whether the issue of endogeneity on the number-of-sibling may alter our conclusion, we estimate equation (2) using instrumental variable approach. The following first-stage equation is estimated:

$$Sib_{ict} = \gamma_0 + \gamma_1 \sum_{n=0}^{2} \omega_n E_{c(t_n)} + \gamma_2 X^P_{i(t_{14})} + \gamma_3 W_{ict} + \gamma_4 Z_{ct_m} + \rho_c + \upsilon_{ict} \tag{3}$$

Table 3-6: Time trends in the CHNS and CFPS

Panel A	CFPS Rural		CHNS Rural		CFPS Rural		CHNS Rural	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time trend	0.142***	0.064***	0.139***	0.128***				
	(0.013)	(0.008)	(0.008)	(0.008)				
Born in 1960s					1.214***	0.762***	1.089**	1.419***
					(0.148)	(0.167)	(0.342)	(0.256)
Born in 1970s					2.290***	1.159***	2.404***	2.371***
					(0.301)	(0.210)	(0.310)	(0.246)
Born in 1980s					4.372***	1.962***	4.411***	4.092***
					(0.419)	(0.247)	(0.249)	(0.220)
Constant	166.386***	157.769***	164.665***	154.268***	167.205***	158.044***	165.434***	154.798***
	(0.245)	(0.139)	(0.138)	(0.133)	(0.162)	(0.076)	(0.203)	(0.148)
Birth region fixed effects	Birth province							
Observations	3219	3077	3707	3341	3219	3077	3707	3341
Adjusted R^2	0.265	0.181	0.195	0.183	0.259	0.178	0.195	0.182

Panel B	CFPS Rural		CHNS Rural		CFPS Rural		CHNS Rural	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time trend	0.143***	0.066***	0.130***	0.121***				
	(0.009)	(0.008)	(0.009)	(0.009)				
Born in 1960s					1.249***	0.752***	1.050***	1.382***
					(0.192)	(0.193)	(0.246)	(0.222)
Born in 1970s					2.275***	1.192***	2.384***	2.319***
					(0.258)	(0.266)	(0.239)	(0.246)
Born in 1980s					4.434***	2.032***	4.130***	3.885***
					(0.292)	(0.254)	(0.330)	(0.292)
Constant	166.372***	157.739***	164.808***	154.375***	167.187***	158.027***	165.491***	154.839***
	(0.157)	(0.150)	(0.158)	(0.144)	(0.137)	(0.135)	(0.143)	(0.134)
Birth region fixed effects	Birth county		Current county		Birth county		Current county	
Observations	3219	3077	3703	3337	3219	3077	3703	3337
Adjusted R^2	0.277	0.196	0.257	0.237	0.271	0.194	0.257	0.238

Note:
1) Robust Standard Errors (SEs) are presented in parentheses;
2) SEs are clustered at the provincial level in Panel A and county level in Panel B for both CFPS and CHNS samples;
3) Provincial fixed effects and county fixed effects are included in Panel A and B, respectively;
4) Individuals born in the CHNS survey provinces are included only;
5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

where t_m indicates individual's birth year, and Z is the instrument. The IV used in this study is a birth-control rate at the county level for each year since the 1950, where birth-control methods include male and female sterilisation, intrauterine devices, pills, use of condoms, and any other method of birth controls. Given that the birth-control rate increased sharply after the introduction of the OCP, if we take the Local Average Treatment Effect (LATE) interpretation of the IV results, the IV estimate of β_4 in equation (2) measures the average sibling effect generated from the increase in the birth-control rate, or generated by the introduction of the OCP.[28]

Our IV is highly correlated with the endogenous variable as indicated in Table 3-A5. We argue that fertile women's share of contraceptive use at the county level should not have direct impact on individuals' adult height over and above its impact through the number-of-siblings variable.

Table 3-7 reports the selected IV results. The F-statistics for the significance of the IV reported at the bottom of the Table indicate the instrument is highly correlated with the sibling numbers. The coefficient on the number of siblings for different samples is still negative and statistically significant, suggesting that the direction of our conclusion using OLS should not be altered significantly. The magnitude of the coefficients, however, increased significantly. This may be related to the LATE interpretation: those whose parents' fertility decision was altered by the change in the birth control rate at the county of resident.

Due to the fact that both our IV and exposure variable have a strong time trend we are unable to estimate the specification with the exposure effect included.

6.3 Missing values in local gazetteer data

As discussed in the data section and detailed in Appendix C, our county-level data have many missing values. We used available years of county-level data on a particular variable together with prefecture/provincial level trend for this variable to predict for the missing years' values.

To check whether our main results are driven by these predicted values we estimate equation (2) excluding all the predicted observations one at a time for the rural sample. Table 3-A6 presents the estimated results for equation (2).[29] These results show that excluding predicted value for missing 'agriculture output' variable, the coefficients on agricultural output slightly increase from 0.746 to 0.768 among males (compare columns 2 and 5) and other coefficients do not change. For the female sample, the coefficient on agricultural output variable decreases from 0.074 to 0.042 (compare columns 3 and 6). For industrial output, the coefficient decreases only by 0.009 for males and increases by 0.054 for females (column 2, 3, 8, and 9). When predicted values of death rates and government social expenditure variables are excluded, the absolute sizes of the coefficients on those two variables increase for both males and females

[28]The second chapter of this thesis provides information on how we collect and construct the birth control rate variable.

[29]We also estimate results with additional 'exposure' variable and its interaction term with the number-of-siblings and the results are largely consistent. These results are available upon request from the authors. Results for urban samples are also available.

Table 3-7: Instrumental variable results for fertility effect

	Rural sample			Urban sample		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Agricultural output p.c.- age 0 to 3	0.045	0.177	-0.101	-0.352**	-0.206	-0.479**
	(0.169)	(0.261)	(0.236)	(0.150)	(0.178)	(0.228)
Log Industrial output p.c.- age 0 to 3	0.087	0.050	0.111	-0.025	-0.166	0.124
	(0.088)	(0.111)	(0.110)	(0.124)	(0.223)	(0.097)
Log Gov. Social expenditure p.c.- age 0 to 3	0.370***	0.464***	0.288**	0.351*	0.560***	0.072
	(0.103)	(0.137)	(0.119)	(0.180)	(0.202)	(0.350)
Death rate (‰) - age 0 to 3	-0.006	0.025	-0.041**	0.107*	0.117*	0.077
	(0.014)	(0.017)	(0.017)	(0.064)	(0.063)	(0.078)
Parental schooling years	0.063***	0.076***	0.044**	0.071*	0.102**	0.082*
	(0.016)	(0.025)	(0.022)	(0.038)	(0.040)	(0.044)
Father's occ rank at child's age 14	0.002***	0.002***	0.002**	0.001	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Father's party membership at child's age 14	0.496***	0.492**	0.447**	-0.146	-0.397	0.061
	(0.137)	(0.204)	(0.194)	(0.298)	(0.383)	(0.332)
Male	9.908***			11.142***		
	(0.094)			(0.221)		
Han Chinese	0.117	0.244	0.004	-0.502	-0.084	-1.033
	(0.248)	(0.355)	(0.327)	(0.949)	(1.067)	(1.014)
Birth order	0.563***	0.814***	0.279*	1.282***	1.269***	1.157*
	(0.111)	(0.171)	(0.151)	(0.405)	(0.444)	(0.663)
Number of siblings	-0.837***	-1.220***	-0.398*	-1.883***	-1.830***	-1.756**
	(0.164)	(0.250)	(0.224)	(0.472)	(0.515)	(0.810)
Dummies for missing values	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16051	8098	7953	2457	1270	1187
Adjusted R^2	0.544	0.166	0.157	0.568	0.138	0.039
Kleibergen-Paap Wald F stat	337.217	242.955	212.332	72.309	78.081	37.191

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples;

3) Unit: log of 10000 yuan; measured in 1990 price level;

4) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values; for father's party membership and occupation are included;

5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(column 2, 3, 11, 12, 14, and 15). Table 3-A7 excludes all the predicted values, which means individuals with at least one predicted value of county-level variables are excluded. Column 4 in Table 3-A7 shows the coefficients on agricultural and industrial output are smaller than in column 1 and insignificant. However, the coefficients for males in column 5 are still sizeable and economically meaningful. Comparing columns 1 to 3 with columns 4 to 6 indicates that the positive government social expenditure effects are not sensitive to the exclusion of all the predicted values. Columns 7 to 9 adds the exposure variable and columns 10 to 12 adds its interaction with the sibling measure.

6.4 Other potential channels

Not only fertility declines but also other channels may have contributed to rising female education performance, such as education policy changes. In this subsection, we discuss potential channels and test whether those drive our main results.

First, China introduced the 9-year compulsory education law in 1986, which requires parents to ensure that their school-age children complete up to junior high school education. The implementation was nationwide and was not divided into urban and rural areas [Liang and Dong, 2019]. Under the law, children under 15 who had already left school had to return to school until they complete 9-years of schooling upon the law's effective date [Fang et al., 2012]. Thus, those born after 1972 (under 15 years old in 1986) are subject to the 9-year compulsory education. If females are less likely to complete 9-years of education than males before the reform, this law could narrow the gender gap in educational attainments (in favour of males), which is highly correlated to cognitive skills. Second, the Chinese government began to rapidly expand tertiary education in 1999 to achieve mass higher education. As a result, higher education enrollment increased by 47% between 1998 and 1999, and it reached 6.3 million in 2009 from 1.0 million in 1998 [Yeung, 2013; Wan, 2006]. It means that the policy significantly expanded higher education opportunities in China. Again, this university expansion may have contributed to reducing the gender inequality in educational attainments if females were underrepresented before the expansion.

A few existing studies suggest relevant findings. The gender gap in education has been narrowed since the 9-year compulsory education law was introduced [McGarry and Sun, 2018], especially in urban areas [Zeng et al., 2014]. With the university expansions, women are less disadvantaged in obtaining higher education [Zhang and Chen, 2014; Yeung, 2013]. However, none of these studies identify the causal relationship between the policy implementations and the gender gap, which could not capture the causal effects.

To test whether the two reforms drive our main findings, we add four variables to the main estimation model in equation 1: (1) 9-year compulsory dummy variable, (2) Univ. expansion dummy variable, (3) an interaction term between the 9-year compulsory variable with the female variable, and (4) an interaction term between the Univ. expansion with the female variable. First, "9-year compulsory" is 1 if one was under 15 (aged between 0 and 14) at

Table 3-8: Other potential channels

	Rural sample				Urban sample			
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Siblings	0.042***	0.126***	0.016**	0.130**	0.034***	0.160	0.023*	0.297
	(0.007)	(0.037)	(0.007)	(0.054)	(0.012)	(0.152)	(0.013)	(0.237)
Siblings*Female	-0.095***	-0.332***	-0.046***	-0.248***	-0.055***	-0.098***	-0.037**	-0.002
	(0.009)	(0.017)	(0.009)	(0.039)	(0.014)	(0.027)	(0.015)	(0.070)
Female	-0.188***	0.060*	-0.436***	-0.101	-0.010	-0.017	-0.075*	-0.126
	(0.026)	(0.035)	(0.030)	(0.078)	(0.029)	(0.032)	(0.038)	(0.081)
9-year compulsory			-0.132***	-0.023			-0.063	-0.016
			(0.025)	(0.033)			(0.043)	(0.054)
9-year compulsory × Female			0.295***	0.090*			0.089*	0.117
			(0.024)	(0.046)			(0.051)	(0.075)
University expansion			-0.104***	-0.075**			-0.132***	-0.122***
			(0.032)	(0.032)			(0.041)	(0.044)
University expansion × Female			0.106***	0.042			0.017	0.038
			(0.028)	(0.029)			(0.039)	(0.047)
Observations	21641	21641	21641	21641	4058	4058	4058	4058
Adjusted R^2	0.410	0.374	0.418	0.397	0.296	0.267	0.297	0.081
Kleibergen-Paap Wald F stat		155.21		76.70		3.43		2.50

Note:
1) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
2) Robust Standard Errors (SEs) are presented in parentheses;
3) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
4) Control variables include the same set of covariates in Table 3 in addition to four variables of the "9-year compulsory" and "University expansion";
5) The instrument is 7-year average Birth Control Rate;
6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

the year of the law introduction and 0 otherwise. Figure 1 shows the law implementation years slightly differ across provinces.[30] Second, "Univ. expansion" is equal to 1 if one started tertiary education after the university expansion in 1999 and 0 otherwise. Those born in 1981 or after were exposed to the expansion as typically students enter universities at 18 years old under the Chinese education system.

Table 3-8 compares our main results (columns 1-2 and 5-6) with the new estimation results controlling for the above-mentioned four variables (columns 3-4 and 7-8). Columns 3 and 4 show that the coefficients on the interaction term between the number of siblings and female dummy are still precisely estimated when the policy change exposure variables are included for the rural sample. The OLS coefficient on the interaction term is also negative and statistically significant for the urban sample (column 7), but the IV coefficient is close to zero (column 8). Considering that Kleibergen-Paap Wald F stat is 2.50, far lower than the rule-of-thumb number 10, the instrument is not strong enough in the urban case.

Overall we show that the education policy changes would not be the main drivers of our findings. Other potential factors may also affect the gender gap in cognitive skills, such as sex ratio or the college entrance exam policy changes over time.[31] However, we cannot test the

[30] If one was born in 1974 in Beijing, where the 9-year compulsory law implemented since 1986, she was subject to the compulsory education because she was 12 years old in 1986.

[31] For example, a highly skewed sex ratio may mean a high competition among males. And such competition may change individual educational choices or parental education investment patterns, which might affect the gender gap in education.

hypotheses in this study, mainly due to limited data availability.

6.5 Other sensitivity tests

As discussed earlier, parental height is an important determinant of a child's height, and our main equation (2) omits this variable due to the limited information available. We observe parental height information only when one's parents also participate in the CFPS survey. Out of 16,051 observations in the rural sample, only 3,141 individuals (19.6%) have mother's height and 2,664 (16.6%) have father's height information. We lose more than 90% of female and about 70% of male observations, and the huge gender difference is possibly due to the fact males are more likely to live with their parents in the same households after marriage than females are. In other words, male parents are more likely to participate in the CFPS survey as they live together. Table 3-A8 compares the estimation results with and without mother's height variable.[32] Column 1 and 4 indicate controlling for parental height would not be a big problem in our analysis. The coefficient on agricultural output decreases from 0.950 to 0.876 but it increases among males (column 2 and 5). The government social expenditure effects become smaller once mother's height is controlled for (column 1, 2, 4, 5) but the coefficients are statistically significant at 5% confidence level and still sizeable. Regardless of the inclusion of mother's height variable, the coefficients on the number of siblings are statistically significant and the sizes do not change much.

Another concern is the potential shrinkage in age 50 to 60 in our sample as discussed in the Data Section. Table 3-A2 shows excluding those over 50 reduces the coefficient size of agricultural output to a large extent (columns 1 and 4), although it is still economically meaningful among males (column 5). Among males, the coefficient on the government social expenditure increase by 69% (columns 2 and 5), and sibling effects on height become almost twice when people over 50 are excluded.

[32] We choose mother's height as it has about 500 more observations. The estimation results with father's height variable are available upon request from the authors.

7 Conclusion

This study explores the relationship between height of individuals born between 1950 and 1990 and economic development and fertility policy changes over the period. We find that China experienced a rapid growth in height of individuals. If we take a linear trend, the rural and urban males decade growth are both above 1 cm (1.3 cm for rural males and 1.5 cm for urban males), which are faster than what was found for the Western European males between 1870 and 1970 (1 cm per decade). The height growth, however, was not linear over the 40-year period, with the fastest growth happening in the 1980s. The the 1980s decade, the height of rural and urban males grew by 2 cm and 2.5 cm, respectively. For rural and urban females the 1980s decade saw 0.9 cm and 1.7 cm growth, respectively.

The average of the 1980s height growth is a combination of two offsetting forces: the introduction of the market-oriented economic reform and the introduction of the OCP. We find that the height growth during this period is mainly driven by the economic reforms which led to fast and sustained improvement in local socioeconomic conditions, whereas the introduction of the OCP deterred height growth for families with a strong son-preference. Controlling for the introduction of OCP, we observe an even larger height growth (2.7 cm for rural and urban males, 1.3 cm for rural females and 2.3 cm for urban females). For the pre-reform era, the single most important factor for height growth is per capita government spending on health and education.

This study provides suggestive evidence as to how the market-oriented economic reforms changed people's lives. In particular, it improved living standards of rural families by boosting agricultural and local industrial output growth which should have directly increased food availability and household income. However, the impact of the OCP in general was negative, especially so for the rural population, and even more so for rural females, predominantly due to the wide-spread son-preference in the rural society.

The Effects of Long-term Single-sex Schooling on Female Academic Achievement

1 Introduction

Many studies have shown that single-sex schooling improves female cognitive performance [Sullivan et al., 2010; Park et al., 2013; Choi et al., 2014; Eisenkopf et al., 2015; Jackson, 2016; Dustmann et al., 2018; Booth et al., 2018b]. It also changes behaviours and preferences [Booth and Nolen, 2012a,b; Booth et al., 2014]. Girls in single-sex environments make more risk-loving and competitive choices than girls in mixed-sex environments. Although these findings conclude that single-sex schooling is beneficial for girls, some evidence suggest its negligible or negative effects [Jackson, 2012; Strain, 2013; Pahlke et al., 2014; Lee et al., 2014; Sohn, 2016]. A cross country meta-analysis finds no significant benefit of single-sex education [Pahlke et al., 2014], primary school students with no opposite gender classmates perform worse at math test than those in mixed-sex classes [Strain, 2013], and the positive coefficient on single-sex school variable becomes negligible when a comprehensive set of the student characteristics are controlled for [Sohn, 2016]. Some studies further explore its heterogeneous effects across different groups [Lee and Lockheed, 1990; Sullivan et al., 2010; Jackson, 2012; Park et al., 2013; Choi et al., 2014; Eisenkopf et al., 2015; Sohn, 2016; Booth et al., 2018b].[1] The benefit of single-sex schooling exists only among students above the median of a test score distribution [Sohn, 2016], and girls who strongly prefer single-sex schooling perform better in single-sex schools than those who do not [Jackson, 2012].

However, the debate over the effectiveness of single-sex education has been on whether it is effective or not rather than how long it can be effective. The latter is important, considering the potential cost of single-sex schooling. Studies in psychology suggest that segregating girls from boys at school highlights gender identity [Halpern et al., 2011; Pahlke et al., 2014; Fabes et al., 2015], which may increase stereotypes and prejudices among students [Bigler

[1] Most studies find that single-sex schooling is positively associated with female academic performance, while the association is not clear among boys [Lee and Lockheed, 1990; Sullivan et al., 2010; Park et al., 2013; Choi et al., 2014; Eisenkopf et al., 2015; Booth et al., 2018b].

and Liben, 2007; Martin and Fabes, 2001; Hilliard and Liben, 2010]. Moreover, some find that gender stereotyping among girls is negatively associated with their math performances [Plante et al., 2013; Igbo et al., 2015].

The main challenges to analyse the impact of long-term exposure to single-sex schooling are as follows. First, it raises endogeneity concerns. Students who choose single-sex education may have different unobservable characteristics (e.g., passion for study) compared to students who do not choose. To overcome this issue, recent studies rely on random school assignments in natural or field experiments. Several studies exploit South Korean government policy that assigns students to middle school and high school within school districts [Park et al., 2013; Choi et al., 2014; Lee et al., 2016; Sohn, 2016]. Other studies randomly assigned students to single-sex classes or groups [Booth and Nolen, 2012a,b; Booth et al., 2018b; Babcock et al., 2017]. However, the solution to the first challenge (endogeneity problem) brings the second challenge - limited coverage of data. The above Korean studies use cross-sectional administrative data on individual test scores. The main reason is it contains school district information, enabling them to control for school district fixed effects and improve internal validity. However, by focusing on the cross-sectional variations in the administrative data, they are unable to see the trajectory of student academic performance moving together with single-sex schooling experience over time. The administrative data provide neither individual experience of single-sex education before high school nor previous academic performance. Studies using field or lab experiments have a comprehensive of individual information, however, the length of exposure to single-sex environments are limited to analyse its long-term exposure effects (e.g., 8 weeks [Booth et al., 2014]; 1 hour per week [Booth et al., 2018a]).

This paper fills in the gap. We examine whether long-term exposure to single-sex schooling is beneficial for female academic development. We attempt to answer this question by 1) exploiting South Korean random school assignment policy that assigning students into single-sex middle and high school and 2) using a panel survey that follows up students form the first-year middle school until high school graduation. Students in our sample attended single-sex middle/high schools from zero up to six years, and we compare academic performance over time. More precisely, we divide students into four groups - students who attend 1) single-sex middle and high school (SSSS), single-sex middle school and mixed-sex high school (SSCO), mixed-sex middle school and single-sex high school (COSS), and mixed-sex middle and high school (COCO) - and compare their test scores over time to see how the relative rankings of performance changes depending on the length of exposure to single-sex schooling.

The results first show the positive effect of attending single-sex middle school on female test scores on Korean, English, and Math during middle school, consistent with existing evidence. Then we also show that girls in single-sex high school perform better than girls in mixed-sex high school at the college entrance exam; however, it only applies to girls who did not attend single-sex middle school. That is, girls exposed to single-sex schooling for the longest period (SSSS) - six years in total - perform worse than girls who experienced single-sex schooling only during high school (COSS) by 17.3%–26.9% of a standard deviation in the

college entrance exam. To put it another way, attending single-sex high schools significantly improves female performance only among those who never experienced single-sex education before high school. It is striking, because if single-sex schooling is beneficial regardless of the exposure length, we expect girls with the longest exposures to perform the best among four groups. The estimates are robust to the inclusion of individual, parental, and school controls. These patterns are not clear among boys.

An important implication of this study is that we show the differential effects of single-sex education within the same individuals over time. Existing studies have documented its heterogenous effects across different demographic groups such as gender [Lee and Bryk, 1986; Park et al., 2013; Dustmann et al., 2018; Booth et al., 2018b] or different subjects [Sohn, 2016; Eisenkopf et al., 2015], however, little is known about the heterogeneity for the same individuals with varying degree of exposure. Our results suggest that single-sex schooling may be or may not be effective depending on how it is designed.

We discuss two potential mechanisms of the adverse effect. First, single-sex schooling may highlight gender identity, resulting in worsening female academic performance. The literature on adolescent development suggests that students in gender-segregated environments are more likely to have gender-stereotype attitudes due to gender salience. And gender stereotyping is negatively associated with female academic performance, such as math. We find that the negative effect of long-term exposure among girls comes from SSSS girls whose high schools had a high proportion of female teachers. That is, among girls who experienced only female classmates during secondary schooling, those highly exposed even to the same gender teachers particularly perform worse than the rest. This pattern supports that single-sex schooling may worsen female cognitive performance by forming gender-salient environments. The second conjecture is that a single-sex environment may trigger intra-sexual competition among girls via gender salience, as suggested by studies on adolescent development. That is, single-sex schooling could result in less cooperative peer interactions by facilitating competition. We find that SSSS girls are less likely to enjoy group assignment and share course material with peers relative to COSS girls.

The rest of this paper is organised as follows. Section 2 provides the background of the education system and random assignment policy in South Korea. Section 3 introduces the data and the main sample for our analysis. Section 4 discusses the empirical strategy. Section 5 presents our main results, Section 6 suggests potential mechanisms, and Section 7 checks the robustness of our results. Section 8 concludes.

2 Background

2.1 Education in South Korea

Formal education in South Korea before college consists of three stages: primary school (grades 1 to 6), middle school (grades 7 to 9), and general/special-purposed/vocational high

school (grades 10 to 12).

Students enter middle school after six years of primary school education, and students are on average 13 years old when they start middle school.[2] Since middle school education is compulsory, all primary school graduates advance to middle schools, and almost all the students attend general middle school.[3] In 2005, 99.5% (703,800) of 707,126 primary school graduates advanced to middle schools, and 99.9% of 703,800 entrants enrolled in general middle schools as shown in Table 4-A1 [KED, 2005].[4] After three years of middle school education, most students advance to high schools. 690,438 students graduated from middle schools in 2008, and 98.7% of them entered high schools [KED, 2008].

Table 4-A2 provides the number of high school entrants in 2008 by school types. There are three major types of high schools in Korea - general, special-purpose, and vocational high schools. 72.3% of the new high school students in 2008 enrolled in general high schools, 25.1% entered vocational high schools, and 2.6% advanced to special-purpose high schools. Students who plan to attend universities in the future choose either general or special-purpose high schools, and those aiming for a job choose vocational education in general. Students with a top-level academic performance during middle school or specialised skills such as art get admitted to special-purpose high schools.

At the end of high school, most students attend the College Scholastic Ability Test (CSAT) - the college entrance exam - to apply for universities unless one wants to get a job with a high school degree. In 2010, 712,227 people attended the CSAT and 541,880 (76.1%) were high school students.[5] We do not have the official information on what proportion of general high school students attend the CSAT. We check the attendance rate among general high school students in survey data - the Korean Education and Employment Panel (KEEP). In 2004, only 0.29% of general high school students surveyed did not attend the CSAT.

2.2 The Equalisation Policy (EP) in Korea

South Korea has a unique education policy which is called "Equalisation Policy (EP)". The policy aims to equalise the quality of education across schools to achieve the equality of education opportunity.

In 1950, the Korean government introduced compulsory education law, which requires children to attend 6-year primary school education [AKS, 2000]. Its implementation delayed during the Korean War and accelerated between 1954 and 1959. Consequently, the number of primary school students increased, leading to high competition among students to enter well-established middle schools. Primary students overworked to prepare the middle school entrance exam, and it raised public concerns over young children's physical and mental

[2]For example, students born between March 1992 and February 1993 started primary school in March 1999, middle school in March 2005, and high school in March 2008 unless they had some special cases.

[3]There are three types of middle schools in Korea - general, physical, and art education middle schools

[4]Students in our main analysis sample entered middle schools in 2005 and high schools in 2008.

[5]Source: Korea Institute of Curriculum & Evaluation; The rest of them are high school graduates (21.7%) and etc (2.0%)

health. The government decided to abolish the middle school entrance exam and randomly assign primary school students into middle schools in 1969 in Seoul. As of 1971, the random assignment policy implemented throughout the nation.

Soon after, the same issue arose in the context of high school education. Before 1974, all high schools set their entrance exams and selected students based on the applicants' exam performance [Kang et al., 2007]. The competition-based admission system induced ranking high schools according to the academic ability of their students. Due to a common perception that one's high school ranking determined the course of one's life [Kang et al., 2007], middle school students struggled with the exam stress, and eventually, the government changed the exam-based admission system to the random assignment-based system since 1974. The EP was introduced in 1974 in Seoul and Busan first and extended to other regions until 1981.

The EP is a package of education policies that target equalisation of the quality of education across general middle/high schools. Each regional education office designs and implements the policy. The first main feature is a random assignment, which aims to equalise the education opportunity for students. Local education offices randomly assign a student into a school within her/his school district that is determined based on where the student lives. Students may move residence to another school district to enter a school they prefer, but they will be subject to another random assignment in the new district [Lee et al., 2014; Park et al., 2013], which means it is not possible to pick one school under the rule.[6] Secondly, the EP attempts to equalise the quality of schools. Under the policy, all private and public schools are publicly funded, follow the unified curriculum, and charge the same tuition fee [Lee et al., 2016; Hahn et al., 2018]. It also controls teacher quality. The government assigns newly hired teachers to public schools regardless of the teachers' preferences, and also, existing teachers have to move to another school every five years within the same school district [Sohn, 2016].

The EP has been controversial mainly due to the possibility of levelling down student quality because the EP mixes high ability students with low ability students, which may lower the overall quality via adverse peer effects [AKS, 2000]. Also, it limits the autonomy of private schools and school choices among students and their parents. Over the last decades since 1974, the local governments had revised, abolished, or reintroduced the EPs. Nowadays, most regions in Korea either partially consider student preferences or implement a preference-based assignment system. Section 3 provides further details of the EP for our study period.

3 Data

This study uses a panel survey of the Korean Education Longitudinal Study (KELS) collected by the Korean Educational Development Institute (KEDI). The KELS surveys 6908 students who started middle school in 2005 and continues to follow up until the students become 30 years old. We use the first six waves, which cover the first-year middle school to the third-year high school until they finish the college entrance exam (CSAT). The KELS randomly

[6] It is possible if the district has only one school. We discuss this issue later in Section 4.

selected nationally representative 150 middle schools using proportional stratified random sampling. It first classified all regions in Korea into four categories - Seoul, 6 metropolitan cities, small cities, and rural areas - and randomly selected middle schools from each category in proportion to the number of middle school students enrolled in each category area. 50 first-year students are randomly selected from each selected middle school. The initially targeted sample includes 6999 students in the first wave, and 6908 among them agreed to participate in the survey.

The KELS has a unique feature compared to other Korean educational panel studies and administrative data. It provides each student's middle school academic ability information, in addition to the college entrance exam (CSAT) scores received at the end of high school. Existing studies on single-sex schooling effects in Korea use administrative data of either individual CSAT [Park et al., 2013; Dustmann et al., 2018] or middle school test performance [Lee et al., 2016], not both, which is unable to discover the trajectory of academic performance over time within the same individuals. The KELS conducted the Middle school Scholastic Ability Test (MSAT) in the first three waves and tested each student's Korean, English, and math abilities.[7] This feature allows us to look at single-sex schooling effects during middle and high school for the same individual over time. We can also estimate the effect conditional on individual ability before the exposure to single-sex schooling experience.[8] The KELS also provides a comprehensive set of student, parents, school, and teacher characteristics.

Table 4-1: Main sample construction

	Observations
Number of students who were surveyed in wave 1	6908
+ Attended general high schools	3916
++ The Equalisation Policy implemented	2404
+++ Living in metropolitan cities	1862
++++ Middle school test score available	1760
+++++ College entrance exam score available	1395
++++++ Parental education/Income info. available	1352

The main sample consists of 1352 students. Table 4-1 shows how we construct the sample. 6908 students agreed to participate in the survey in the first wave. Since the EP does not apply to non-general middle/high schools, we only include 3916 students who attended general middle/high schools out of 6908. Among them, 2404 students started middle/high schools (i.e., in 2005 and 2008) in regions where the EP was being implemented.[9] We further exclude 542 (2404-1862=542) students living in non-metropolitan regions to ensure internal validity of the estimation results. In other words, our sample covers metropolitan regions

[7]The Korean Educational Development Institute designed the MSAT to evaluate middle school students' comprehensive scholastic abilities. The Korean exam consists of 35 questions, English exam 33 questions, and Math exam 20 questions. More information are available upon request.

[8]The KELS also asked each respondent regarding their primary school academic performance, and the respondents answered by choosing one number between 1 to 9 (very poor 1 to very good 9). We control for this variable in the main analysis.

[9]Each school administrator reported whether the school region was subject to the EP in each wave.

of South Korea.[10] The reasons are as follows: First, we cannot identify small cities in non-metropolitan sample, but the main estimation strategy is to include city fixed effects. That is, we exclude non-metropolitan observations because we cannot control for city fixed effects. Second, columns (A)-(C) in Table 4-A3 show that each school district has less than 2 middle schools in non-metropolitan areas, which possibly enables students to pick a school by moving residence to a district.[11] Lastly, column (a)-(c) in Table 4-A3 indicates many school districts are not subject to the EP (e.g., 30 middle school districts are not under the EP in Gyeonggi-do.), although the case is rare in metropolitan cities. The first outcome variable in the main analysis is the third-year middle school test (MSAT) score, and 1760 of 1862 individuals have that information. The second outcome variable is the college entrance exam (CSAT) score, and we lose 365 (20.7%) of 1760 observations due to missing information. All these 365 students refused to report the CSAT scores, although they participated in the survey at the year. Unfortunately, we are unable to identify the reason why they refused to provide the CSAT results. Alternatively, Figure 4-A1 compares the MSAT scores of students who are included in the sample with those excluded. It indicates that students included have higher MSAT Korean scores than those excluded (compare the thin solid line with the solid line), which is due to excluding students in non-general high school students (mainly vocational high school students; compare the thin solid line with the short-dashed line).

Table 4-2 summarises the main sample by gender of students. It consists of 636 female and 716 male students.[12] Panel A indicates that 32-33% of students in our sample attended single-sex middle schools. 67% of female students and 70% of male students studied in single-sex high schools.

Panel B shows that female students perform substantially better than male students at Korean and English tests, while male students perform as good as or better than females at Math tests. We standardised the MSAT scores with zero mean and one standard deviation among all 6283 test-takers. The CSAT scores are standardised with zero mean and one standard deviation among 3854 students whose scores are available in the KELS raw data. Section check shows later that our findings are not sensitive to the standardisation methods. The average MSAT and CSAT z-scores are above zeros in the main sample, and Figure 4-A1 shows that it is mainly because we exclude vocational high school students who have low academic performances relative to academic high school students.

Panel C presents the school average of middle school characteristics. About 80% of middle schools are public schools. On average, teachers have completed over 16 years of education(equivalent to a 4-year university degree) and worked for 13.7 years. 70-71% of teachers in middle schools are female. The annual budgets on student learning activities are 175700 won in the female sample and 157400 won in the male sample (Approximately 140-150 US dollars).

[10] Figure 4-A7 shows the geographic coverage of our main sample. According to the Korean Education Statistical yearbook, 681,444 students entered high schools in Korea, and 46% of them were in metropolitan cities.

[11] We discuss further in Section 4.1

[12] The proportion of males (52.96%) in our analysis sample is not different from that of the initial KELS sample. 52.37% of 6908 respondents are male.

Table 4-2: Summary statistics

	Female		Male	
	Mean	SD	Mean	SD
Panel A				
Middle school: Single-sex[1)]	0.33	(0.47)	0.32	(0.47)
High school: Single-sex[1)]	0.67	(0.47)	0.70	(0.46)
Panel B				
MSAT standardized score - Korean	0.58	(0.79)	0.12	(0.91)
MSAT standardized score - English	0.61	(0.85)	0.17	(0.97)
MSAT standardized score - Math	0.30	(0.93)	0.30	(0.96)
CSAT standardized score - Korean	0.11	(0.94)	-0.10	(1.04)
CSAT standardized score - English	0.09	(0.95)	-0.08	(1.04)
CSAT standardized score - Math	-0.05	(0.99)	0.05	(1.00)
Panel C				
Middle school: Public school	0.79	(0.40)	0.80	(0.40)
Middle school: Class size	36.44	(4.15)	36.38	(3.76)
Middle school: Teacher schooling years	16.59	(0.24)	16.60	(0.25)
Middle school: Teacher experience years	13.79	(3.31)	13.69	(4.07)
Middle school: Female teacher ratio	0.71	(0.15)	0.70	(0.16)
Middle school: Budget per student	17.57	(9.90)	15.74	(9.14)
Middle school: % students entered general high school	0.73	(0.08)	0.73	(0.07)
Panel D				
High school: Public school	0.50	(0.50)	0.37	(0.48)
High school: Class size	37.09	(2.82)	35.85	(3.65)
High school: Teacher schooling years	16.75	(0.24)	16.82	(0.24)
High school: Teacher experience years	17.10	(3.00)	17.81	(3.17)
High school: Female teacher ratio	0.53	(0.14)	0.34	(0.20)
Panel E				
Age in 2010 (months)	219.41	(3.70)	219.70	(3.57)
Household monthly income (10 thousands won)	387.27	(222.26)	425.01	(425.68)
Parental mean schooling years	13.38	(1.93)	13.33	(2.19)
Private education experience in middle school	0.77	(0.42)	0.82	(0.38)
Primary school academic performance	6.50	(1.63)	6.48	(1.74)
Observations	636		716	

Note:
1) The female sample includes 9 single-sex middle schools and 45 mixed-sex middle schools. The male sample includes 10 single-sex middle schools and 45 mixed-sex middle schools;
2) The female sample includes 50 single-sex high schools and 39 mixed-sex high schools. The male sample includes 52 single-sex high schools and 38 mixed-sex high schools.

The proportion of students who advanced to general high school captures the overall school students' academic abilities because a low proportion means relatively many students go to vocational high schools.

Panel D provides high school characteristics information. 50% and 37% of high schools are publicly owned in the female and male samples, respectively. About 36-37 students are in per class, teachers have completed more than 16 years of education, and teachers have 17 years of experience. 53% of high school teachers are female for the female sample and only 34% of teachers are female for the male sample.

Panel E indicates students are 18 years old on average in the third year of high school (in 2010). The monthly household income is 387 thousands won (approximately 3250 US dollars) for females and 425 thousands won among males.[13] On average, parents have completed over 13 years of education, which is equivalent to an additional year of education after high school graduation. The KELS asked each student to report their primary school performance on a scale of 1 (lowest) to 9 (highest), and the average is 6.5.

Table 4-3: Number of students by four group in the main sample

	Female sample		Male sample	
	# students	%	# students	%
Single-sex middle school - Single-sex high school (SSSS)	155	24.4%	172	24.0%
Single-sex middle school - Mixed-sex high school (SSCO)	58	9.1%	58	8.1%
Mixed-sex middle school - Single-sex high school (COSS)	268	42.1%	330	46.1%
Mixed-sex middle school - Mixed-sex high school (COCO)	155	24.4%	156	21.8%
Total	636	100%	716	100%

Our main analysis divides students into four groups based on duration of single-sex schooling exposure and timing - 1) students who attended single-sex middle and high school (SSSS), 2) students who attended single-sex middle school and mixed-sex high school (SSCO), 3) students who attended mixed-sex middle school and single-sex high school (COSS), and 4) students who attended mixed-sex middle and high school (COCO). Table 4-3 shows the sample distribution. 24.4% of students (155 students) in our sample went to single-sex middle school and high school. Less than 10% of students attended single-sex middle schools and mixed-sex high schools. COSS group accounts 42.1%-46.1% for the total observations, and 21.8%-24.4% of students attended mixed-sex middle and high schools.

[13]The large gender difference comes from 4 outliers in the male sample. Excluding the outliers, the average of household monthly income for males is 402.65 (10 thousands won), and the gender differences is not statistically significant at 10% significance level.

4 Empirical strategy

We estimate the different effects of single-sex schooling across groups defined by different duration of exposure to single-sex schooling. To do so, we compare the academic outcomes of the 4 group by estimating equation (4-1) and (4-2). First, equation (4-1) estimates the effect of attending single-sex middle schools on student middle school test (MSAT) scores achieved in the third year of middle school.

$$Y_{imhcs} = \alpha_0 + \alpha_1 SSSS_{mh} + \alpha_2 SSCO_{mh} + \alpha_3 COCO_{mh} + \alpha_4 PSAT_i + \sum_{k=1}^{7} \gamma_k M_{mk} + \sum_{k=1}^{5} \lambda_k X_{ki} + \delta_c + \epsilon_{imhcs} \tag{4-1}$$

$$s \in \{Korean, English, Math\}$$

where Y_{imhcs} represents the MSAT z-score on subject s of a student i who attended a middle school m and a high school h located in a metropolitan city c. The test subject s includes Korean, English, and Math.[14] The four group indicates - $SSSS_{mh}$, $SSCO_{mh}$, and $COCO_{mh}$ - are included, and $COSS_{mh}$ is the omitted reference category. Students did not start high school when they attended the MSAT, so α_1 and α_2 both capture single-sex schooling effects during middle school. We control for 7 middle school characteristics M_{mk} presented in Table 4-2 - public middle school dummy, the average class size, the average teachers' schooling years, the average teachers' experience years, female teacher ratio, the annual budget on learning activities per student, and the proportion of students who entered general high school.[15]. We also include individual and household characteristics X_{ki} - log household income, mean parental schooling years, and private education experience -, and self-reported primary school academic performance $PSAT_i$ is included. δ_c captures city fixed effects. Next, equation (4-2) estimates the effect of attending single-sex middle and high schools on the college entrance exam (CSAT) outcomes achieved in the end of high school.

$$\begin{aligned} Y_{imhcs} = {} & \beta_0 + \beta_1 SSSS_{mh} + \beta_2 SSCO_{mh} + \beta_3 COCO_{mh} + \beta_4 MSAT_{ims} \\ & + \beta_5 PSAT_i + \sum_{k=1}^{7} \alpha_k M_{mk} + \sum_{k=1}^{5} \eta_k H_{hk} + \sum_{k=1}^{5} \tau_k X_{ki} + \zeta_c + \epsilon_{imhc} \end{aligned} \tag{4-2}$$

$$s \in \{Korean, English, Math\}$$

where Y_{imhcs} represents the CSAT z-score on subject s of a student i who attended a middle school m and high school h located in a metropolitan city c. We include the same set of the three group indicators as in equation (4-1). We control for 7 middle school variables M_{mk} and 5 high school variables H_{hk} - public high school dummy, the average class size, the average teachers' schooling years, the average teachers' experience years, and female teacher ratio. We again include the same set of individual and household characteristics X_{ki}. The self-reported

[14] The KELS conducted the MSAT for these three subjects only.
[15] The last two variables are available only for middle school

primary school academic performance $PSAT_i$ and the middle school test scores $MSAT_{ims}$ for subject s are added to consider basic academic abilities.

α_1, α_2, and α_3 capture the mean difference in the MSAT z-scores and β_1, β_2, and β_3 capture the mean difference in the CSAT z-scores between SSSS and COSS groups, SSCO and COSS groups, and COCO and COSS groups, respectively. However, if students can choose single-sex schooling, αs and βs would be biased. Suppose students with a higher level of passion for study than others are more likely to choose a single-sex middle school, then we expect α_1 and α_2 to be upward biased and α_3 to be downward biased. Furthermore, students who already experienced single-sex middle school may have systematically different single-sex schooling choice behaviours for high school. If one finds single-sex schooling helpful, he/she will choose a single-sex high school, then β_1 would be upward biased unless we control for individual intrinsic preference over single-sex schooling environments, which are unobservable.

4.1 Endogeneity issue

To resolve the endogeneity concerns, we exploit that students are randomly assigned to single-sex middle and high school in Korea, as discussed in Section 2. However, several issues threaten our identification strategy. In this subsection, we provide the details of the EP in our study period (i.e., in 2005 for middle school assignment and 2008 for high school assignment) and discuss potential problems in our estimates.

First, the random assignment was implemented within a school district in 2005 and 2008, but we do not have school district information. The KELS only provides city-level identification data; in other words, we can only control for δ_c instead of district fixed effects. Thus we cannot control for unobservables varying across school districts within cities. We assume that the covariates in equation (4-1) and (4-2) can address heterogeneous variations in the omitted variables affecting single-sex schooling choice and academic performance across school districts within a city.

Second, the EP implementation does not always mean the full randomisation assignment. Regional education offices can decide whether to consider student preferences over schools partially or not [Lee et al., 2014]. We phone-interviewed each city and local education office in the metropolitan cities in July 2016.[16] It includes education offices in Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, and Ulsan. Table 4-4 reports the outcomes. When the officer refused to answer or did not know the details, we assume the policy was the same as that of other regions within the city.[17] The assumption is reasonable because local education offices tend to unify the EP rules within the same city.[18] In 2005, 4 of 7 metropolitan cities assigned

[16]Each metropolitan city has a city education office and many local offices. The city office is in charge of high school education, including implementing the random assignment policy, and the local offices administer primary school and middle school education. For example, Seoul has one city education office and 11 local offices that are in charge of each local area education matters.

[17]In Korea, the government moves officers to another department every a few to several years. Some of the officers claimed that information were not available because they were not in the department in either 2005 or 2008.

[18]There are some exceptions. Two of all the metropolitan cities had such exceptions. One is Seoul (See Sohn

students into middle schools on a full random basis within districts. 3 out of 7 metropolitan cities partially considered student preference over middle schools in 2005. Suppose a local education office considers student preferences by 40% of the total enrolment. It means 40% of new students for each school will be those who rank the school first.[19] To fill the rest (60%), students who are not assigned yet will be randomly assigned to schools within its school district.[20] In 2008, among seven metropolitan cities, only Seoul implemented the full random assignment policy. Five metropolitan cities partially considered student preferences by 40%-60% of the total enrolments, and Incheon considered student preference and middle school academic performance for the assignment implementation. Students are first assigned to high schools based on their school rankings. When a school has an excess demand, the education office ranks students by individual middle school academic performances and give priority to those with better performances. Our KELS data does not allow us to consider each assignment rule due to limited information. We discuss this issue more in Section 7 and check the robustness of our results.

Table 4-4: Random assignment details by cities

	Middle school assignment in 2005	High school assignment in 2008
Seoul	Random	Random
Busan	Random	Partial random
Incheon	Random	Partial random
Daejeon	Random	Partial random
Daegu	40% preference + 60% random	Partial random
Gwangju	Partial random	Partial random
Ulsan	Partial random	Partial random

Source: City and local education offices phone interviews

Third, the randomisation assignment policy does not help us identify the causal effect if some school districts have only either single-sex or mixed-sex schools. It is because students can choose one of the types by simply moving house to the district. We cannot check whether every district has at least one single-sex school and mixed-sex school in our data as the KELS provides neither school district information nor its location. Alternatively, Table 4-A3 shows the average number of schools per district.[21] Column (A), (B), and (C) in Panel A shows Seoul has 6.3 mixed-sex middle schools, 1 girls' middle school, and 1 boys' middle school on average, which means that every school district has at least one mixed-sex middle school

[2016] for details) and the other is Gwangju city. According to an officer in Gwangju, they partially considered student preferences when assigning them to middle schools in all school districts except for one district due to a special circumstances.

[19] If the number of students who rank the school first is lower than 40% of the school's total enrolment, students who rank the school second also can be assigned to the school. If the number of students who rank the school first is higher than 40% of the school's total enrolment, the local education office randomly chooses students among them.

[20] The local education offices did not provide more detailed information on how they implemented and what percentage of students were assigned to their preferred schools.

[21] We can find this information only for 2016 onwards. According to local education officers, school district division had rarely changed over the last decade.

on average. Students who prefer single-sex schools would fail to find a school district only with single-sex schools, which means moving residence might not be the solution in this case. If one prefers a mixed-sex middle school, he/she may move to the school district without single-sex schools. Because we find the positive impact of single-sex middle school, it is not the major concern. The number of mixed-sex middle schools per district in most non-metropolitan areas is only around 1. Also, the average number of middle schools per district is less than 2 in most regions. It suggests that the random assignment policy may not work in a way that is supposed to work in non-metropolitan areas due to the limited number of schools per district. Column (a), (b), (f), and (g) indicates that non-trivial number of school districts are not subject to the EP in non-metropolitan areas, while almost all metropolitan cities have no non-EP school districts.

In conclusion, our identification strategy relies on the within-district random assignment policy in metropolitan regions, but we cannot successfully exploit the random variations due to data limitation. We discuss this issue more throughout the paper and suggest potential solutions at the end for future research.

4.2 Balance check

In this subsection, we check differences in characteristics across the four groups. The methodology is to regress each variable on the three group dummy variables - SSSS, SSCO, and COCO - using COSS as the reference category. Table 4-5 presents the balance test results.

Column 1 in Panel A and B uses public middle school dummy as a dependent variable. The coefficient -0.210 indicates that middle schools in the SSSS group are 21%p less likely to be publicly owned than middle schools in the COSS group controlling for city fixed effects. The coefficient sizes are large, and the coefficient on SSSS and SSCO dummies in Panel B are statistically significant at 5% significance level. Single-sex schools are more likely to be private schools in Korea, but, as mentioned earlier, public and private schools are equally funded, use the unified curriculums, and charge the same tuition fee. However, private schools have autonomy in teacher hiring decisions, and the teachers have a higher incentive to perform well than those in public schools [Hahn et al., 2018]. We see similar patterns in column 8 where public high school dummy is the dependent variable. In Section 7. Robustness check, we include public schools only and test test if the main results are sensitive to the school type.

Most of the differences in characteristics between SSSS and COSS girls are neither statistically significant nor big enough to have economic meanings except for the middle school budget on learning activities per student in column 7. The coefficient 6.335 says that middle schools that SSSS girls attended spent 63350 Korean won (roughly equivalent to 50 US dollars) more on learning activities per student than schools that COSS girls attended within the same city. In contrast, this difference is not significant among boys.[22] Among males, SSSS

[22]Later the main results show the positive effect of attending single-sex middle school among girls. It might be driven by the difference in school budget, although we control for this variable. However, if that is the case, we should see the pattern among students in COCO groups, and we do not.

boys have 2.807 smaller class sizes, 0.121 years less educated teachers, and 5.5% more female teachers in high school, but the magnitude of the differences are not substantial. Column 13 to 17 show that no significant differences are detected in terms of individual and household characteristics between SSSS and COSS groups.

The rest of the columns show statistically significant differences across the four groups; however, the differences are not substantial to be economically meaningful in most cases.

Table 4-5: Balance test

	Middle school characteristics							High school characteristics					Individual characteristics				
Panel A: Female	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
SS-SS	-0.210	-0.211	0.085	0.216	-0.058	-0.032	6.335*	-0.033	-1.663**	-0.073	-0.685	0.007	-0.277	1.211	-0.055	0.480	-0.250
	(0.160)	(0.954)	(0.096)	(0.857)	(0.055)	(0.026)	(3.267)	(0.123)	(0.797)	(0.057)	(0.537)	(0.035)	(0.311)	(25.741)	(0.049)	(0.452)	(0.168)
SS-CO	-0.332	-0.482	0.054	1.711***	-0.007	0.006	1.428	0.151	-0.308	0.199***	0.715	-0.095***	-0.538	-11.423	0.002	0.675	0.315
	(0.233)	(1.116)	(0.108)	(0.607)	(0.036)	(0.029)	(5.361)	(0.093)	(1.218)	(0.045)	(0.586)	(0.031)	(0.345)	(41.059)	(0.061)	(0.604)	(0.213)
CO-CO	0.031	0.033	0.052	-0.723	-0.010	0.015	4.797**	0.270***	-0.100	0.100**	-0.679	-0.043	-0.216	38.674	0.029	0.818**	-0.134
	(0.088)	(0.829)	(0.049)	(0.615)	(0.032)	(0.014)	(1.984)	(0.095)	(0.505)	(0.047)	(0.539)	(0.033)	(0.232)	(34.605)	(0.050)	(0.381)	(0.185)
Obs.	636	636	636	636	636	636	636	636	636	636	636	636	636	636	636	636	636
Adjusted R^2	0.222	0.599	0.258	0.355	0.234	0.379	0.195	0.255	0.209	0.216	0.356	0.241	0.040	0.041	0.020	0.003	0.023
Control mean	0.87	36.36	16.58	13.93	0.70	0.74	14.89	0.19	36.41	16.84	18.43	0.25	13.58	406.13	0.84	219.59	6.54
	Middle school characteristics							High school characteristics					Individual characteristics				
Panel B: Male	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
SS-SS	-0.424**	-0.335	0.088	0.085	-0.089	-0.027	-0.324	0.151**	-2.807***	-0.121**	0.015	0.055**	-0.245	106.137	-0.055	-0.218	-0.119
	(0.170)	(1.000)	(0.109)	(1.229)	(0.079)	(0.029)	(2.155)	(0.058)	(0.845)	(0.053)	(0.447)	(0.024)	(0.313)	(65.555)	(0.038)	(0.323)	(0.174)
SS-CO	-0.382**	-0.969	0.152	0.144	-0.080	-0.039	-0.042	0.330**	-3.765**	0.019	-1.233**	0.197***	-0.824*	-216.605	-0.054	0.466	-0.680**
	(0.159)	(0.803)	(0.169)	(1.593)	(0.076)	(0.027)	(2.799)	(0.145)	(1.518)	(0.051)	(0.603)	(0.059)	(0.477)	(186.314)	(0.052)	(0.484)	(0.260)
CO-CO	-0.079	-0.058	0.026	-0.203	-0.007	0.016	5.176*	0.453***	0.739	-0.016	-0.988	0.218***	0.024	-31.885	-0.008	0.582	-0.058
	(0.075)	(0.671)	(0.051)	(0.707)	(0.023)	(0.015)	(2.785)	(0.102)	(0.460)	(0.040)	(0.739)	(0.032)	(0.270)	(27.094)	(0.044)	(0.357)	(0.130)
Obs.	716	716	716	716	716	716	716	716	716	716	716	716	716	716	716	716	716
AdjustedR^2	0.224	0.525	0.060	0.426	0.228	0.190	0.118	0.214	0.217	0.168	0.198	0.305	0.041	0.037	-0.006	-0.003	0.009
Control mean	0.79	36.68	16.59	13.77	0.69	0.73	15.59	0.34	37.47	16.70	17.75	0.52	13.70	379.75	0.80	219.07	6.59

Dependent variables

(1) Public school dummy (2) Average class size (3) Average teacher schooling years (4) Average teacher experience years (5) Female teacher ratio
(6) % students who advanced to general high schools (7) Average budget on learning activities per student (8) Public school dummy (9) Average class size
(10) Average teacher schooling years (11) Average teacher experience years (12) Average female teacher ratio (13) Parental mean schooling years
(14) Log household income (15) Private tutoring experience in the 3rd year middle school (16) Age in months (17) Primary school academic performance

Note:

1) Robust standard errors clustered by middle school;
2) City fixed effects are included in all columns;
3) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

5 Results

5.1 Unconditional mean comparisons

Figure 4-1 below compares middle school academic performance - MSAT Korean - between students in single-sex and mixed-sex schools. The KELS conducted this test at the end of the third-year middle school (in November 2007), so students in single-sex middle schools had been exposed to single-sex schooling for about three years. Girls in single-sex middle schools perform better at the MSAT than girls in mixed-sex middle schools, which is consistent with existing evidence on the positive impact of single-sex schooling on female academic performance[Sullivan et al., 2010; Park et al., 2013; Eisenkopf et al., 2015; Sohn, 2016; Jackson, 2016; Dustmann et al., 2018; Booth et al., 2018b]. Figure 4-A2 indicates single-sex school girls also achieved higher scores at the English and Math tests. The difference is relatively smaller among boys. Figure 4-1 (Right) shows boys in single-sex middle schools also perform better than those in mixed-sex middle schools, but the difference is not clear in the right tail of the distribution. Moreover, Figure 4-A2 shows no significant difference in English and Math performances among boys.

Figure 4-1: Middle school performance comparison - (Left) Girls (Right) Boys

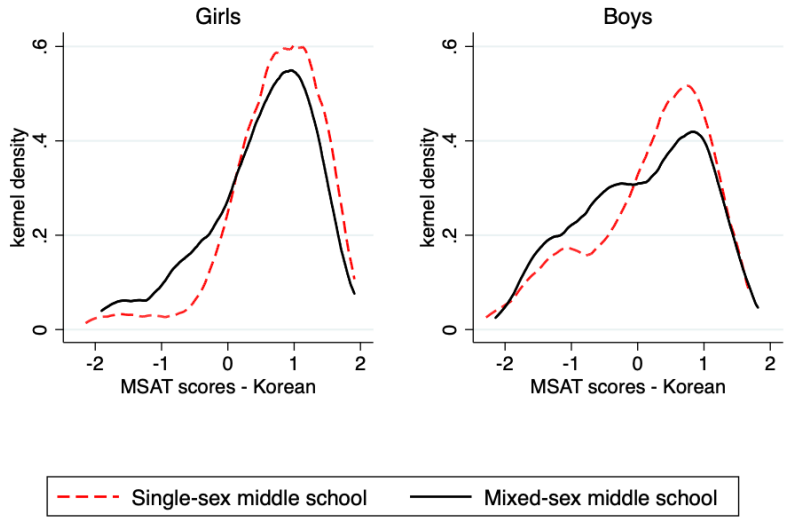
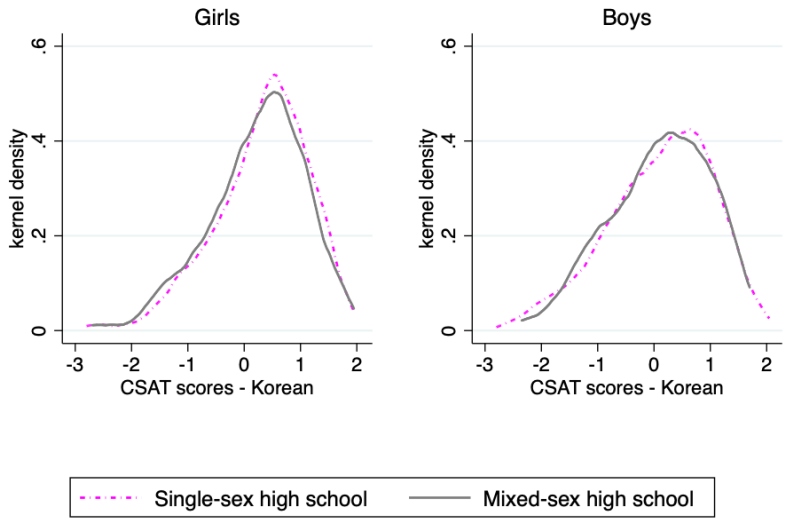


Figure 4-2 compares student performance in the college entrance exam (CSAT) carried out in November 2010 - the third year in high school. Girls in single-sex high schools perform slightly better at the CSAT Korean than girls in mixed-sex high schools, and Figure 4-A3 shows the difference is more distinct in the CSAT English and Math performances. Studies that look at the single-sex high school effects on the CSAT outcomes in South Korea also find a positive impact among girls [Park et al., 2013; Sohn, 2016; Dustmann et al., 2018]. Among boys, the difference is not as significant as that for girls as far as the CSAT Korean performance. Figure 4-A3 suggests that boys in single-sex high schools perform slightly better

than those in mixed-sex high schools in the English and Math tests.

Figure 4-2: High school performance comparison - (Left) Girls (Right) Boys

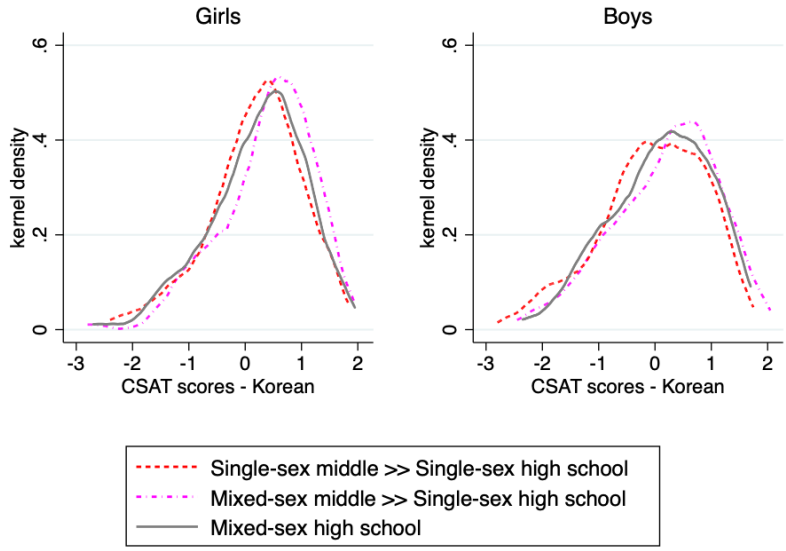


However, students in single-sex high schools consist of two groups - 1) students who experienced single-sex schooling for three years in middle school and another three years in high school (six-year exposure in total; SSSS) and 2) students who experienced single-sex schooling for the first time in high school (three-year exposure in total; COSS). Previous literature examining the single-sex schooling effect during high school does not consider this fact due to limited information. For example, studies on single-sex schooling effect in Korea use only either individual CSAT test score data with high school information [Park et al., 2013; Dustmann et al., 2018; Sohn, 2016] or individual national-level test score data with middle school information [Lee et al., 2016]. Figure 4-3 further divides single-sex high school students into two groups - SSSS and COSS.[23] Figure 4-3 (Left) shows a clear difference in the CSAT Korean score distributions between SSSS and COSS girls. Better outcomes in single-sex high schools shown in Figure 4-2 come from COSS girls exposed to single-sex schooling for three years. SSSS girls spent 6-year of the entire secondary schooling period in single-sex environments, and their performance distribution is quite close to that of students in mixed-sex high schools. The patterns are not as clear among boys. Figure 4-A4 indicates the pattern is even clearer in the Math outcomes.

In summary, single-sex education seems to improve girls' academic outcomes during middle school, consistent with existing evidence. Also, single-sex schooling seems to improve female academic performance during high school, considering that COSS girls achieved higher than those in mixed-sex high schools at the CSAT. However, it does not seem to be effective for those girls who already experienced single-sex environments since the first year in middle

[23]For simplicity of the figure, Figure 4-3 does not further divide mixed-sex high school students. Later we use the four group divisions in the main analysis.

Figure 4-3: High school performance comparison - (Left) Girls (Right) Boys



school - SSSS girls. Figure 4-A5 shows that SSSS girls outperformed girls in mixed-sex middle schools. That is, it is not likely that pre-high school ability differences drive this pattern.

5.2 Main results

Since the above figures compare unconditional mean of test scores, this section provides the estimation results that capture the mean differences conditional on individual, household, and school characteristics as mentioned in equation (4-1) and (4-2).

Table 4-6 presents the result. The dependent variables are individual middle school standardised test scores in the third year of middle school (MSAT) in column 1 to 3 and college entrance exam standardised test scores in the third year of high school (CSAT) in column 4 to 6.

Columns 1 to 3 in Panel A indicate that girls in single-sex middle schools (SSSS and SSCO girls) performed significantly better at the MSAT than girls in mixed-sex middle schools. SSSS girls achieved 16.5%-27.8% of a standard deviation higher scores than the reference category of COSS girls during middle school. The differences are statistically significant at 5% significance level. SSCO girls also performed as good as SSSS girls relative to COSS girls. On the other hand, the differences between the two mixed-sex middle school groups (COSS and COCO girls) are not statistically significant. Overall it suggests the positive effect of attending single-sex middle schools on their middle school academic performances, which is consistent with literature evidence on its positive impacts among female students [Sullivan et al., 2010; Park et al., 2013; Eisenkopf et al., 2015; Sohn, 2016; Jackson, 2016; Dustmann et al., 2018; Booth et al., 2018b].

The pattern is not clear among boys. Columns 1 to 3 in Panel B show that the MSAT performances are not statistically different across the four groups among boys. Furthermore,

Table 4-6: Main results

Panel A: Girls	Middle school exam (MSAT)			College entrance exam (CSAT)		
	(1) Korean	(2) English	(3) Math	(4) Korean	(5) English	(6) Math
SS-SS	0.278***	0.165**	0.217**	-0.173***	-0.179***	-0.269***
	(0.088)	(0.075)	(0.090)	(0.060)	(0.062)	(0.074)
SS-CO	0.237***	0.245***	0.221**	0.037	-0.175*	-0.359***
	(0.088)	(0.072)	(0.106)	(0.084)	(0.089)	(0.103)
CO-CO	0.116	-0.056	0.095	-0.065	-0.088	-0.288***
	(0.071)	(0.079)	(0.117)	(0.075)	(0.066)	(0.101)
Observations	636	636	636	636	636	636
Adjusted R^2	0.213	0.293	0.291	0.326	0.449	0.350

Panel B: Boys	Middle school exam (MSAT)			College entrance exam (CSAT)		
	(1) Korean	(2) English	(3) Math	(4) Korean	(5) English	(6) Math
SS-SS	0.051	0.117	-0.080	-0.076	-0.102	-0.083
	(0.122)	(0.109)	(0.107)	(0.103)	(0.108)	(0.128)
SS-CO	-0.023	-0.059	-0.047	0.057	-0.009	-0.094
	(0.111)	(0.124)	(0.137)	(0.096)	(0.100)	(0.122)
CO-CO	0.043	0.012	0.040	0.094	-0.007	-0.190*
	(0.087)	(0.098)	(0.118)	(0.079)	(0.110)	(0.111)
Observations	716	716	716	716	716	716
Adjusted R^2	0.070	0.151	0.170	0.284	0.331	0.186

Note:
1) The dependent variables are the MSAT z-scores in column (1)-(3) and the CSAT z-scores in column (4)-(6)
2) Column (1)-(3) control for middle school characteristics (public school dummy, teacher schooling years, class size, teacher experience years, female teacher ratio, the percentage of students who advanced to general high school, middle school budget for learning activities per student), log of household monthly income, parental mean schooling years, student age in months, primary school academic performance level (1 to 9), and city fixed effects;
3) Column (4)-(6) include all the covariates in column (1)-(3) in addition to high school characteristics (public school dummy, teacher schooling years, class size, teacher experience years, female teacher ratio)
4) Robust standard errors in parentheses are clustered by middle school;
5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the estimated coefficient sizes are not as big as that of girls. It is not surprising considering that many existing studies also find that single-sex environments change female performance or behaviours, while it does not have substantial impacts among males [Lee and Lockheed, 1990; Sullivan et al., 2010; Park et al., 2013; Eisenkopf et al., 2015; Booth et al., 2018b]. For example, Booth et al. [2018b] show that female university students who attended all-female classes perform better than female students in mixed-sex classes, but the difference is not significant among male students.

We also use the first and second year MSAT z-scores as dependent variables and estimate the effects in equation (4-1). Table 4-A9 shows girls in single-sex middle school consistently perform better than girls in mixed-sex middle school in the first three years, while the patterns are not clear among boys in 4-A10.

Columns 4 to 6 in Table 4-6 present the differences in the CSAT performance three years later, at the end of high school education. Panel A indicates that girls in mixed-sex high schools (SSCO and COCO) perform worse than the reference category of COSS girls at the CSAT, especially in Math, which shows the positive single-sex high school effects on their academic performance at the end of high school, again consistent with literature findings [Sullivan et al., 2010; Park et al., 2013; Eisenkopf et al., 2015; Sohn, 2016; Jackson, 2016; Dustmann et al., 2018; Booth et al., 2018b]. COCO girls achieve 6.5%-28.8% of a standard deviation lower CSAT scores, and SSCO girls perform worse by 17.5%-35.9% of a standard deviation compared to COSS girls.

Considering our results and existing evidence, we expect SSSS girls to perform the best among four groups as those have been exposed to single-sex schooling the most. However, Panel A in Table 4-6 indicates that it is not true. SSSS girls perform significantly worse than COSS girls by 17.3%-26.9% of a standard deviation. Furthermore, SSSS girls even did worse than COCO girls at the CSAT Korean and English; the coefficients -0.173 and -0.179 are statistically different from the coefficients 0.065 and -0.088 at 5% significance level, respectively.

An important implication here is that we show the heterogeneous effects of single-sex education within the same individuals over time. Existing studies have documented differential effects of single-sex schooling across different demographic groups such as gender [Lee and Bryk, 1986; Park et al., 2013; Dustmann et al., 2018; Booth et al., 2018b] or different subjects [Sohn, 2016; Eisenkopf et al., 2015], however, little is known about the heterogeneity for the same individuals with varying degree of exposure. Our results suggest that single-sex schooling may be or may not be effective depending on how it is designed.

Table 4-6 Panel B shows almost none of the coefficients are statistically significant at 10% significance level. SSSS boys perform worse than COSS boys across all the three subjects, but the differences are not substantial and not as large as that of the female group. Table 4-A8 includes students from non-metropolitan regions under the EP and shows the consistent patterns as our main results in 4-6.

In summary, we find that single-sex schooling improves female academic performance after the first three year of exposure in both middle and high school periods. However, girls

who spend her entire secondary schooling period only in single-sex schooling environments perform as bad as students without single-sex schooling experiences. This suggests not only single-sex schooling but also the duration of its exposure matters. Those effects are negligible among boys.

6 Mechanism

What would possibly drive the adverse effects of long-term exposure to single-sex schooling? This subsection discusses potential mechanisms.

6.1 Highlighting gender

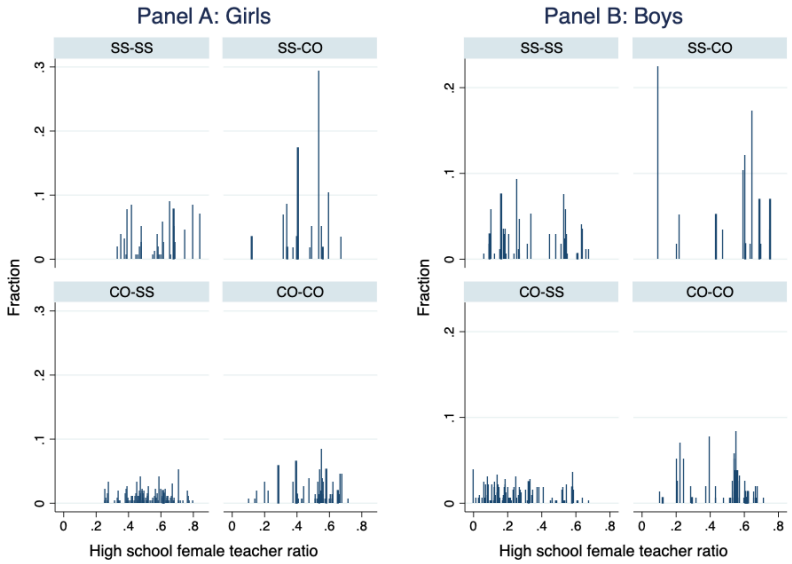
Literature in psychology suggests that single-sex schooling may highlight gender stereotypes [Halpern et al., 2011; Pahlke et al., 2014; Fabes et al., 2015]. In particular, according to the development intergroup theory, separating individuals by their characteristics (e.g., gender or race) makes the difference more salient, which increases stereotypes and prejudices among students [Bigler and Liben, 2007; Martin and Fabes, 2001; Hilliard and Liben, 2010]. It means that segregating students by gender may increase gender stereotypes among them. For example, Hilliard and Liben [2010] conducted an experiment to test how gender salience environments could affect children's gender attitudes in a preschool. Teachers in the treatment group used gender-specific language (labelling groups and individuals with gender terms) and gender-based organisation in the classroom (e.g., assigning boys and girls to separate bulletin boards, lining up boys and girls separately for recess). After the two weeks of experience, children in the treatment group showed significantly increased gender stereotypes and decreased play with other-sex peers. If such gender salience in single-sex environments is negatively associated with female student academic performance, it could explain our results. In fact, some studies suggest that the level of gender stereotyping is negatively associated with female math performances [Plante et al., 2013; Igbo et al., 2015].

To check this mechanism, we test whether the adverse effect of long-term exposure to single-sex schooling comes from schools with a relatively higher degree of gender salience. That is, we need variable capturing gender salience environmental factors. Since the KELS did not ask any relevant questions, instead, we use a proxy variable 'female teacher ratio' as the measure of gender salience environment. We assume schools (particularly single-sex schools) with a higher female teacher ratio have more gender salience environments.

The empirical strategy is to generate an interaction term between the high school female teacher ratio and each group dummy and estimate equation (4-2).[24] Figure 4-4 plots the high school female teacher ratio across the four groups by gender. It ranges 0 percent to 84.21 percent. All-female high schools (SSSS and COSS) have a higher female teacher ratio (Panel A), and all-male high schools have lower female teacher ratio (Panel B) relative to mixed-sex

[24] In other words, we add three interaction terms - SSSS*Female teacher ratio, SSCO*Female teacher ratio, and COCO*Female teacher ratio - to equation 4-2

Figure 4-4: High school Female teacher ratio



high schools. For example, more than 80% of teachers are female in some high schools that SSSS girls attend (Panel A Top-left). More than 60% of teachers are female among COCO boys, while the majority of high schools have more male teachers (i.e., female teacher ratio is less than 50%) among COSS boys (Panel B Bottom-left).

Column 1 in Table 4-7 shows that as the female teacher ratio increases by 1%p, the CSAT Korean scores increase by 0.704% of a standard deviation among COSS girls. Although the magnitudes are different, an increase in female teacher ratio is positively associated with COCO and SSCO girls' CSAT Korean performances, consistent with previous findings [Carrell et al., 2010; Lim and Meer, 2017].[25]

However, an increase in female teacher ratio lowers academic outcomes among SSSS girls. Column 1 shows a 1%p increase in female teacher ratio in high schools decreases SSSS girls' CSAT Korean scores by 0.480% (-1.184+0.704) of a standard deviation. More importantly, once the interaction terms are included, the SSSS coefficients become positive (0.556, 0.710, 0.351), and the magnitudes are larger than SSCO and COCO coefficients (-0.332, 0.236, -0.224; -0.049, -0.088, -0.221). That is to say, girls exposed to the longest period of single-sex education achieved the best scores among the four groups. However, as the female teacher ratio raises, the positive impact decreases and become negative at a certain point. We check the point at which SSSS girls' performance become lower than that of COSS girls in Figure 4-A6. When the high school female teacher ratio is zero, SSSS girls perform the best among the four groups

[25] According to existing studies, teacher gender matters to the students. Carrell et al. (2010) shows that female professors increase female student academic performance at the university, while it does not seem to be influential among male students [Carrell et al., 2010]. Korean female middle school students do significantly better at academic tests when they are taught by female teachers [Lim and Meer, 2017]. However, only few studies discover the effect of teacher-student gender matching in single-sex environments. Lee et al. (2016) shows teachers may change teaching disciplines depending on gender of the students. According to their mechanism analysis, teachers in all-male school tend to choose stricter teaching disciplines, which leads students to study harder [Lee et al., 2016].

Table 4-7: Female teacher ratio

	Girls			Boys		
	(1) Korean	(2) English	(3) Math	(4) Korean	(5) English	(6) Math
SSSS	0.556**	0.710***	0.351	-0.109	-0.176	-0.152
	(0.248)	(0.248)	(0.360)	(0.131)	(0.176)	(0.240)
SSCO	-0.332	0.236	-0.224	0.050	0.042	-0.178
	(0.445)	(0.262)	(0.540)	(0.209)	(0.172)	(0.236)
COCO	-0.049	-0.088	-0.221	-0.147	-0.314	-0.175
	(0.239)	(0.213)	(0.332)	(0.207)	(0.218)	(0.246)
SSSS × High school Female teacher ratio	-1.184**	-1.463***	-0.993	0.137	0.265	0.214
	(0.464)	(0.448)	(0.634)	(0.327)	(0.360)	(0.449)
SSCO × High school Female teacher ratio	0.822	-0.739	-0.198	0.099	0.008	0.195
	(0.921)	(0.466)	(0.985)	(0.389)	(0.409)	(0.451)
COCO × High school Female teacher ratio	0.067	0.100	-0.027	0.583	0.750	0.001
	(0.479)	(0.401)	(0.582)	(0.484)	(0.460)	(0.490)
High school Female teacher ratio	0.704*	0.726	-0.170	-0.601	-0.514	-0.228
	(0.408)	(0.475)	(0.499)	(0.410)	(0.461)	(0.534)
Observations	636	636	636	716	716	716
Adjusted R^2	0.395	0.491	0.382	0.354	0.362	0.218

Note:
1) The dependent variables are the CSAT z-scores;
2) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

as captured by the coefficient 0.556 in column 1 of Table 4-7. SSSS girls' CSAT Korean scores go down as the female teacher ratio increases, while other girls' performance increase.[26] The four lines cross when the ratio reaches 0.5, which means that in a single-sex school where there are more female teachers than male teachers, SSSS girls perform worse than COSS girls. Table 4-A4 tests if such heterogeneous effects existed during middle school and shows no statistically significant differences in the female teacher ratio effects across the four groups. It seems to suggest that a high female teacher proportion is negatively associated with girls' outcomes only when exposed to single-sex education environments for a long time, which leads to a high degree of gender salient environments.

6.2 Peer interactions

The next question is then how highlighting gender adversely affects female academic performance. Some studies on adolescent development argue that single-sex environments trigger intra-sexual competition among girls due to gender salience, while boys do not change behaviours in all-male situations [Lereya et al., 2014]. Similar findings are available in economics literature. Girls in single-sex schools are more likely to choose to compete [Booth and Nolen, 2012a] and prefer real-stakes lottery over a sure bet than girls in mixed-sex environments [Booth et al., 2014], girls are more likely to volunteer than boys but not in all-female environments [Babcock et al., 2017], and such differences not significant among boys. While the behavioural changes in single-sex environment facilitate female achievement in general, according to Lereya et al. [2014], such tensions among girls may result in a bullying environment at school. That is, girls exposed to long-term single-sex schooling may have different peer interactive behaviours at school, which is associated with their academic performance.

We test whether peer interaction in single-sex schools can explain our findings. We use individual self-reported information on her/his interaction with peers and class environments. The KELS asked the same set of attitudinal questions in the third year of middle school and high school, and we use the questions to see changes over time using the main specification in 4-2. The questions include 1) *"I help classmates who fall behind with school work"*, 2) *"I discuss with my classmates to understand difficult class material*, 3) *"I share my class notes or material with my classmates*, and 4) *"I enjoy working on group assignments, which are mainly about one's peer interaction in class*. Each respondent chooses 1 (Strongly disagree) to 5 (Strongly agree), and we convert the variables to an indicator of agreeing (equal to 1) or not (equal to 0).

Table 4-A5 presents the results. Column 2 shows that SSSS girls are more likely to discuss with classmates to understand difficult material in middle school by 3.5%p; however, SSSS girls are less likely to do so in high school by 12.4%p although both SSSS and COSS girls are in single-sex high schools. However, the probability of interacting with peers among SSCO and COCO are also lower than that of COSS girls (compare column 1 with 5, 2 with 6, 3 with

[26] We compute the slopes using the estimated coefficients on the female teacher ratio for each group in column 1 of Table 4-7. The slope is -0.480 (=-1.184+0.704) for SSSS girls, 0.704 for COSS girls, 1.526 (=0.822+0.704) for SSCO girls, and 0.771 (=0.067+0.704) for COCO girls.

7, and 4 with 8). These findings seem to suggest that 1) girls in single-sex high schools interact with peers for learning but 2) girls exposed to single-sex environments for a long time tend to have passive attitudes. Consistent with our main findings, it does not indicate whether single-sex schooling is better or not. It means that single-sex school education requires the right design due to the potential adverse effects.

7 Robustness check

Public and Private schools

The main sample includes students from both private and public middle/high schools, and private schools are more likely to be single-sex. 30% of single-sex middle schools are private, while only 16% of mixed-sex middle schools are private. 66% of single-sex high schools are private, and only 36% of mixed-sex high schools are private. One may argue that single-sex schooling effects may confounded with private schooling effects such as the school quality. In Korea, however, both private and public schools are publicly funded, follow the national curriculum, and charge the same amount of tuition fee since EP introduction [Lee et al., 2016; Hahn et al., 2018]. The only difference is that private schools have autonomy on hiring decision [Hahn et al., 2018]. Teachers in public schools are employed by the government, whereas the private school directors hire teachers in private schools. Table 4-A6 only includes public high schools and shows the patterns are consistent with our main findings.

Early admissions

Universities in Korea offer early admissions based on student academic records in high school before the CSAT. This early decision system is called "Soosi". Students who target the Soosi admissions tend to spend more time on achieving high scores in their high school exams and focus less on the CSAT because some of Soosi applications either do not require the CSAT scores or consider the CSAT scores to a lesser extent compared to those who target the regular admissions. In general, students who have consistently achieved good academic grades during high school prepare for the Soosi application. If SSSS girls have better test scores and are more likely to target Soosi application, SSSS girls would have lower CSAT scores, not because of poor performance. To consider that we control for two variables - 1) a variable indicating if one plans to apply for Soosi admission and 2) a variable indicating if one focuses more on preparing Soosi admission application than regular admission. Table 4-A7 shows it does not significantly change the main results.

Randomness check

Our identification strategy relies on the EP to overcome endogeneity problem, but most metropolitan cities in our sample partially considered student preferences (See Table 4-4). To see whether attending single-sex schools is systematically related to student characteristics, we estimate the probability of entering single-sex middle/high schools as follows:

$$y_i^* = x_i'\beta + u_i \text{ and } y_i = 1[y_i^*\beta + u_i \geq \tau]$$
$$x_i = (1, x_{i2}, ..., x_{ik})' \text{ and } \beta = (\beta_1, ..., \beta_k)'$$

where y_i^* is a probability of entering single-sex school. If the probability of being assigned to a single-school exceeds a certain point τ, a student i is assigned to a single-sex school and we observe $y_i = 1$ which indicates whether one attended a single-sex school. Independent variables x_i include a set of covariates that may affect individual single-sex schooling choice such as household income, parental education level, and primary school academic performance. As discussed earlier, students may move places to another school district to change the probability of being assigned to a single-sex school if one has a strong preference. We include the years of living in the current place as an explanatory variable to capture that possibility.

Table 4-8 presents the Probit estimation results. When it comes to single-sex middle school assignment in column 1 and 2, none of the coefficients are statistically significant at 10% level.[27] For example, an additional year of living in the current place is associated with 2.1%p decrease in entering single-sex school for girls and 0.5%p increase for boys. An additional year of parental schooling decreases the probability of being assigned to single-sex schools among girls by 4.6%p and among boys by 5.9%p.

Columns 3 and 4 shows the result on the probability of attending single-sex high school. We additionally include a dummy variable of single-sex middle school attendance and the MSAT z-score variables. An additional year of parental schooling is associated with an increase in the probability of entering single-sex high school by 6.2%p. Overall most of the estimated coefficients are not statistically significant at 10% levels, and the magnitudes are not economically meaningful. Although it cannot suggest that the school assignments were random, this result at least lessens concerns over systematic differences in student characteristics between single-sex and mixed-sex schools.

[27] One city has only mixed-sex middle schools so it is excluded from the estimation as its city dummy perfectly correlated with the probability

Table 4-8: Randomness check: Single-sex middle school entrance

	Middle school		High school	
	Girls (1)	Boys (2)	Girls (3)	Boys (4)
Years of living in the current residence	-0.021	0.005	0.012	0.010
	(0.014)	(0.011)	(0.012)	(0.010)
Primary school academic performance 1 to 9	0.006	-0.032	-0.010	0.035
	(0.036)	(0.035)	(0.039)	(0.033)
Parental schooling years	-0.048	-0.059	0.062*	-0.033
	(0.036)	(0.039)	(0.036)	(0.030)
Log household income	-0.110	0.064	-0.172	0.270***
	(0.128)	(0.102)	(0.176)	(0.101)
Student age	0.011	-0.010	-0.026*	-0.029*
	(0.014)	(0.013)	(0.015)	(0.016)
MSAT Korean z-score			-0.089	-0.032
			(0.077)	(0.079)
MSAT English z-score			0.088	0.107
			(0.089)	(0.075)
MSAT Math z-score			-0.114	-0.191**
			(0.112)	(0.089)
Observations	573	571	631	705
Pseudo R^2	0.119	0.198	0.108	0.173

Note:
1) The dependent variables are single-sex middle school attendance in columns 1 and 2 and single-sex high school attendance in columns 3 and 4;
2) In addition to the covariates presented, we include city fixed effects;
3) Robust standard errors in parentheses are clustered by middle school;
3) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Conclusion and future research

While recent studies have shown that single-sex education improves female academic outcomes, little is known as to how long it can be effective. This paper explores this question by exploiting the random school assignment policy in South Korea. Consistent with existing evidence, we find that attending single-sex schooling improves female academic performance; however, it depends on the duration of exposure. First, girls in single-sex middle school perform better than girls in mixed-sex middle school at academic tests at the end of middle school - after three years of exposure to single-sex schooling. Second, girls in single-sex high school also perform better than girls in mixed-sex high school at the college entrance exam - after three years of exposure. However, girls who spent 6-year of the entire secondary school period in single-sex environments perform as low as girls in mixed-sex high school at the college entrance exam. The estimates are robust to the inclusion of controls of individual, parental, and school characteristics. These patterns are not clear among boys.

This study has two key limitations: First, the identification strategy relies on the random school assignment, however, the assignment partially considered student preferences in most of our sample regions. Second, we control for city fixed effects instead of school district effects, although the randomisation was done within school districts due to data availability.

Despite these limitations, to our best knowledge, this is the first paper exploring long-term exposure to single-sex schooling under the random assignment policy setting and suggesting the potential adverse effects of long-term single-sex education. Existing studies mainly look at its short-term effects, and also, none of those studies could look at the trajectory of single-sex schooling effect on academic outcomes over time. Further research is required to demonstrate the potential adverse effects of single-sex schooling by tracking student academic performance and behavioural changes over time.

Conclusion

This dissertation shows the importance of childhood environments on cognitive and health development in adulthood. It contributes to the vast literature to explore causal relationships by exploiting natural experiments and introducing unique sets of data on childhood environments. Each paper is independent, but there are common themes that they share.

The first theme relates to gender, more specifically, son-preference culture. Chapter 2 shows that an additional sibling worsens female cognitive skills more than for males in China, where son-preference is prevalent. It results in widening the gender gap in cognitive skills in adulthood in favour of men. The magnitude of female disadvantage is bigger in regions where more people have son preference. Chapter 3 finds the positive and strong association between local economic environments in early childhood and physical condition in adulthood, and the association is strong and significant among males, while it is weaker among females. In particular, the benefit of economic well-being improvement from China's market-oriented reform substantially increases male height, but the effects are weaker among females. All these findings link to son-preference culture in those countries that have practiced Confucian traditions. Sons play the major roles in families and are expected to be the breadwinners, which incentivises parents to allocate more resources to sons than daughters. In fact, Chapter 2 provides suggestive evidence that son-preference culture is the main driver of widening the gender gap. Female disadvantages in big families are more prevalent in regions where (1) people prefer sons over daughters, (2) people believe that women should give birth to at least one boy to continue the family lineage. Chapter 4 shows that separating girls from boys in school significantly changes female academic performance in South Korea, while it is not effective among boys. This suggests the gender gap in education and other human capital related areas can be narrowed by school environment changes.

Another theme is the role of government policies implemented in childhood on one's lifelong outcomes. China implemented strict family-planning policies since the early 1970s due to concerns over food shortage. Chapter 2 suggests that a dramatic decline in fertility introduced by the policy led to a fast improvement in cognitive skills and education. Chapter 3 reveals the importance of local government social expenditures and economic conditions in early childhood. Between 1950 to 1990, Chinese average adult height has increased by more than 1 cm a decade, which is an impressive achievement. We find this growth in height is mainly explained by local government spending on social services, including healthcare,

when individuals are between ages 0 and 3. The market-oriented reforms introduced in the late 1970s also led to speedy growth in height among the 1980s birth cohorts, possibly by expanding farmers' economic opportunities and incomes. Chapter 4 adds a relevant point: an education policy of segregating students by gender could be a solution to improve female cognitive ability even after high school when it is designed properly.

One interesting question that arises in this dissertation is why a strict fertility control narrowed the gender gap in cognitive skills while widening the gender gap in height. A conjecture is that the strict and long-lasting birth control policies may lower parents' expectations of having a son as time goes by, which increases investments in their current daughters gradually over time. That is, to save resources for the future son, parents may not invest in daughters enough in their early childhood but gradually increase investments as they give up a son as time goes by. Lack of investments for daughters in early childhood may widen the gender gap in height because 40% of individual height is determined between age 0 and age 3. Belated investments for daughters in later childhood may not affect height but significantly would improve daughters' cognitive skills. Future research needs to closely look in to how parental son-preference related behaviours or attitudes interact with the fertility control policy over time.

Nonetheless, this dissertation greatly improves our knowledge about childhood environmental effects on adult outcomes, particularly in son-preference cultures.

Appendix to Chapter 2

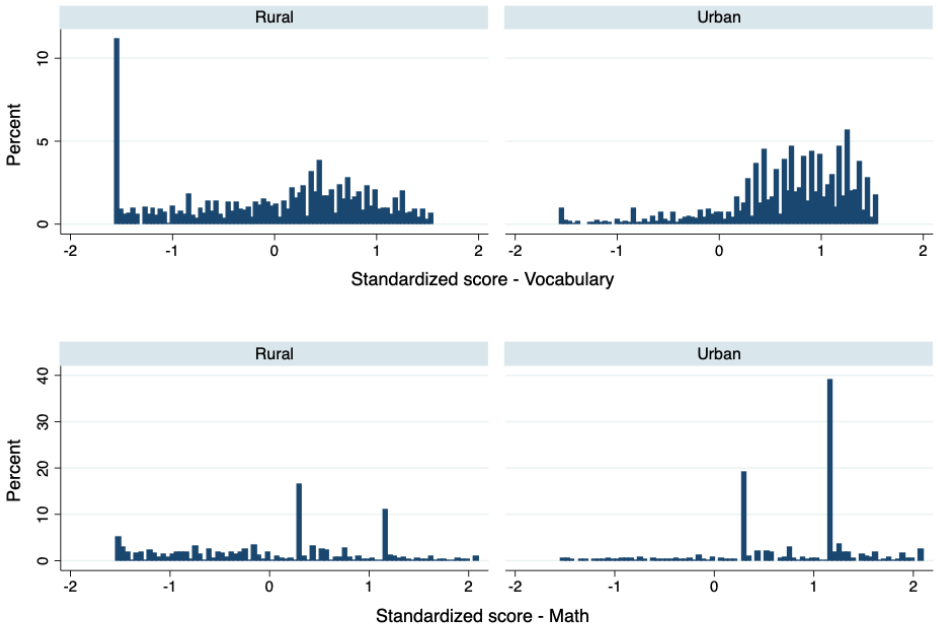
APPENDIX A: Cognitive ability measures in the CFPS

The China Family Panel Studies (CFPS) test respondents aged over 10 on their cognitive abilities. In 2010 and 2014, the CFPS conducted (1) vocabulary test and (2) math test. The vocabulary test consists of 34 Chinese characters drawn from the language textbooks used in primary and secondary schools. The test seeks to measure one's vocabulary skill by checking whether one correctly recognises those characters. These 34 questions are in an ascending order of difficulty, in other words, the 34th question is the hardest. There are 8 sets of 34 questions available and a computer randomly selects a set for each respondent. All the 8 sets are design to have the same difficulty level.[1] The next step is to assign each respondent to a question to begin with. To conduct more efficiently, each respondent is assigned to one of three educational groups - Group 1) primary school graduation or below, Group 2) junior middle school graduation, and Group 3) senior middle school graduation or above. Group 1 starts with the easiest question (i.e., Question 1), Group 2 starts from Question 9, and Group 3 starts from Question 21. The Chinese characters are presented in cards and one is asked to read them aloud. The final score is the number of correct answers including easier questions that were not given to one. For example, if one is from Group 3 and answers to Question 21, Question 22, and Question 23 correctly, the final score is 23. The math test consists of 24 questions including addition, subtraction, multiplication, division, exponents, logarithms, trigonometric functions, sequence, permutation and combination, etc.

A problem is that the CFPS changed test procedure slightly between 2010 and 2014 and we use the combined scores. In 2014, when an interviewer shows the first character (i.e., the first question for one's education group), if one does not have a correct answer, her next question is the first question for the last education group. For example, if one is in Group 3, then she starts from Question 21. If she does not know the right answer for Question 21, her next question is Question 9, which is the first question for a lower education group - Group 2. If she answers Question 9 correctly then continue to solve until she fails to correctly answer 3 questions in order. On the other hand, in 2010, one has no chance to try any question for lower education group. In the above example, if she fails to answer Question 21, her next question is Question 22 and Question 23. If she fails to answer these three in a row, her test is over, and

[1] Unfortunately, the CFPS does not provide the question set.

Figure 2-A1: Main outcome variable - standardised vocabulary and math test scores



her final score is 20. This means that the 2010 test procedure assumes that she can answer all the questions easier than her first question considering her education level. In other words, the lowest possible score for her is 0 in 2014, while it is 20 in 2010.

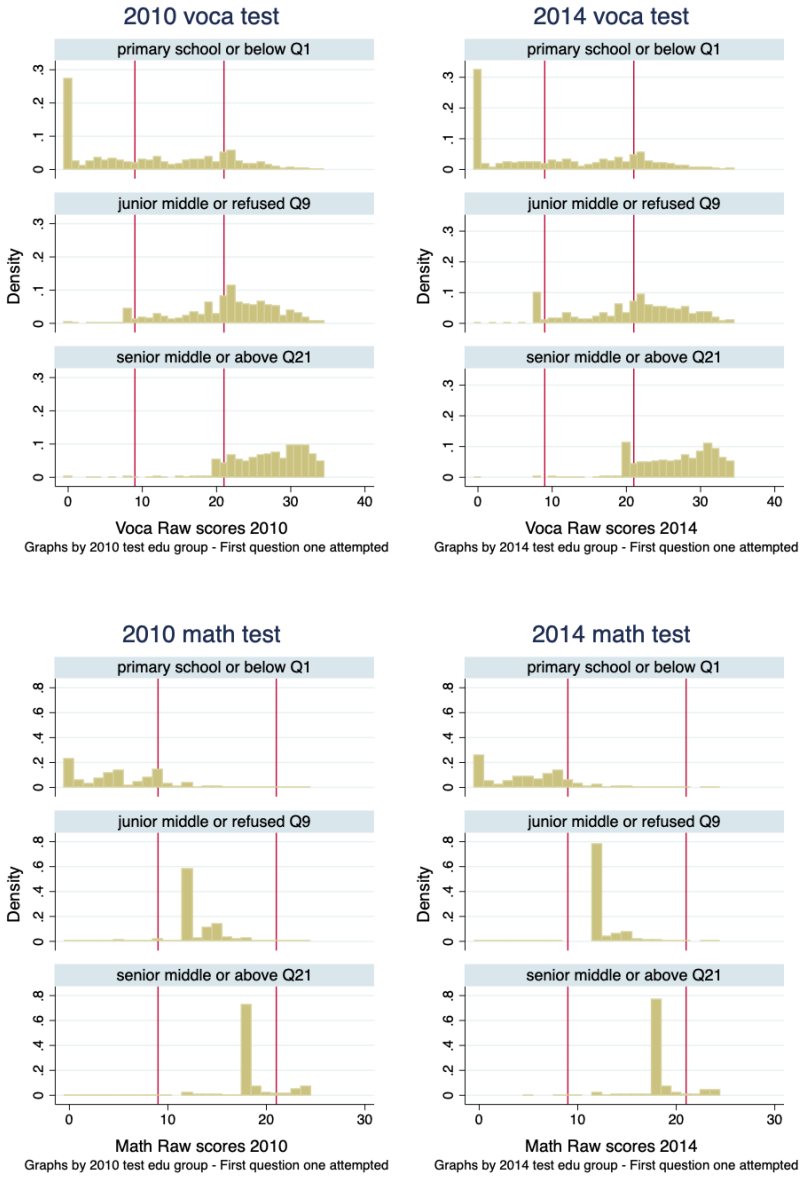
In 2014, the CFPS updated the procedure slightly. In 2010, if one cannot read the first 3 characters in a row or one solves the 34th character, the test ends. On the other hand, in 2014, if one does not have a correct answer to her first question, her next question is the first question for one step lower education group. For example, if one completed senior high school education then she starts with Question 21 and is asked in order. If she cannot answer this first question, her next question is the first question for one step below education group - Question 9 - in 2014. On the other hand, in 2010, she is asked Question 21, 22, and 23 in order regardless of her answers. If she fails to answer all the three questions, her test ends then final score is 20 by assuming that she could solve the first 20 questions considering her education level. In 2014, her final score would be 12 if she could solve up to only Question 12.

Suppose $Q_{1st,g}$ is the first question number for group g and $g = 1, 2, 3$. If one answers $Q_{1st,g}$ correctly and fails to answer the next three questions, her final score is $Q_{1st,g}$ regardless of test year. If one fails to answer the first three questions including $Q_{1st,g}$, in 2010, her final score is $Q_{1st,g} - 1$ by assuming that she can solve all the questions up to $Q_{1st,g} - 1$ considering her education level. On the other hand, in 2014, if one fails to answer the first question $Q_{1st,g}$, her next question is $Q_{1st,k-1}$ and she continues until she gets three questions incorrectly in a row.

In summary, the 2010 test assumes that one can correctly answer all questions that are easier than her group's first question considering her education level. If this assumption is wrong, there should be significant numbers of people whose final scores are lower than

$Q_{1st,k} - 1$ in 2014. If this happens, we are not able to combine the 2010 and 2014 scores together. Figure 2-A2 compares the distributions between 2010 and 2014 tests for each education group. For only few people, the final scores are lower than their $Q_{1st,k} - 1$. For example, among 'senior middle graduation or above group' people, we do not see a meaningful number of people who got below 20 in 2010 and 2014 both. In other words, most of the respondents who failed to answer their first question correctly answered up to $Q_{1st,k} - 1$. In conclusion, it is unlikely that the slight difference in test procedure affects individual scores significantly so we combine test scores from 2010 and 2014.

Figure 2-A2: Test score distribution by test year and education group



APPENDIX B: Main sample selection process

Table 2-B1 shows how we construct the main sample. The CFPS surveyed 33,600 respondents in 2010 and 37,147 in 2014. There are 50,435 individuals surveyed in either or both of the two years. About 81.5% of 50,435 respondents have information on the household registration - *hukou* - status at age 3. 6,093 individuals had urban *hukou* and 34,991 people had rural *hukou* at their age 3 in the combined sample.

The urban sample originally has 6,903 individual observations and about 94.9% of those have test scores available. We focus on 4,515 people born between 1950 and 1990 and 4,070 out of them have sibling information including the number of siblings and birth order. The final sample size is 4,058 excluding observations without birth county information. The rural sample has 34,991 individuals and 93.7% of them have test scores available. The final sample size is 21,641.

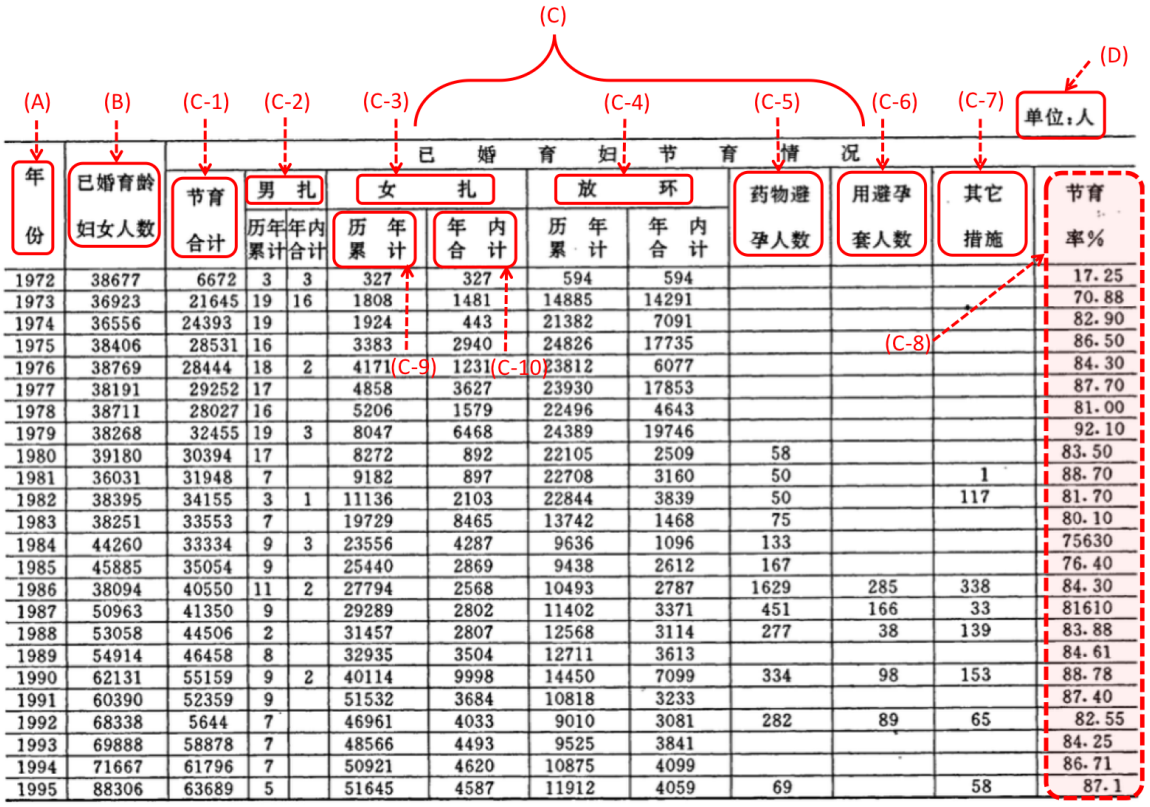
Table 2-B1: Main working sample construction

(Unit: Individual)	2010 sample [1]	2014 tracked [2]	2014 refresh [3]	working sample [1]+[3]
		Total sample		
Original sample	33600	25618	11529	45129
hukou info. at 3 available	33361	25441	6142	39503
With cognitive test	33337	25433	5225	38562
Born 1950-1990	25102	19501	4036	29138
Sibling info. available	24847	19329	1785	26632
Birth order info. available	24284	18928	1542	25826
With non-missing other covariates	24281	18926	1418	25699
		Rural *hukou* sample		
Original sample	28193	22221	5461	33,654
With cognitive test	28183	22220	4601	32,784
Born 1950-1990	21070	16992	3553	24,623
Sibling info. available	20878	16858	1552	22,430
Birth order info. available	20410	16512	1346	21,756
With non-missing other covariates	20408	16511	1233	21,641
		Urban *hukou* sample		
Original sample	5168	3220	681	5,849
With cognitive test	5154	3213	624	5,778
Born 1950-1990	4032	2509	483	4,515
Sibling info. available	3969	2471	233	4,202
Birth order info. available	3874	2416	196	4,070
With non-missing other covariates	3873	2415	185	4,058

APPENDIX C: Birth Control Rate (BCR) data from local gazetteers

We collect the government's birth control policy information from each county/district/prefecture gazetteers. Although each provides information in its own format and covers different periods, there are common variables reported. Figure 2-C1 provides an example of a typical data format. The Birth Control Rate (BCR) is in column (C-8). It is calculated as dividing the number of fertile and married women who use birth control (C-1) by the total number of fertile and married women (B). Note that (C-1) is the number of women who use birth control in the corresponding year, not the number of women who start birth control in that year. For example, column (C-3) indicates 9182 women are sterile in 1981, and 897 of them had sterilisation operation (new cases) in that year. The total number of women using contraception is 7+9182+22708+50+1=31948, and the BCR is 31948/36031=88.668091%. When full information are not available in county gazetteers we additionally check its prefecture or provincial gazetteers and statistical yearbooks. The main sources are (1) county or prefecture level general gazetteers, (2) county or prefecture level health gazetteers, (3) county or prefecture level fertility control gazetteers, (4) municipal or provincial population statistical yearbooks, and (5) municipal or provincial fertility control statistical compilations.

Figure 2-C1: Birth Control Rate in a county's gazetteer



年份	已婚育龄妇女人数	节育合计	男扎 历年累计	男扎 年内合计	女扎 历年累计	女扎 年内合计	放环 历年累计	放环 年内合计	药物避孕人数	用避孕套人数	其它措施	节育率%
1972	38677	6672	3	3	327	327	594	594				17.25
1973	36923	21645	19	16	1808	1481	14885	14291				70.88
1974	36556	24393	19		1924	443	21382	7091				82.90
1975	38406	28531	16		3383	2940	24826	17735				86.50
1976	38769	28444	18	2	4171	1231	23812	6077				84.30
1977	38191	29252	17		4858	3627	23930	17853				87.70
1978	38711	28027	16		5206	1579	22496	4643				81.00
1979	38268	32455	19	3	8047	6468	24389	19746				92.10
1980	39180	30394	17		8272	892	22105	2509	58			83.50
1981	36031	31948	7		9182	897	22708	3160	50		1	88.70
1982	38395	34155	3	1	11136	2103	22844	3839	50		117	81.70
1983	38251	33553	7		19729	8465	13742	1468	75			80.10
1984	44260	33334	9	3	23556	4287	9636	1096	133			75630
1985	45885	35054	9		25440	2869	9438	2612	167			76.40
1986	38094	40550	11	2	27794	2568	10493	2787	1629	285	338	84.30
1987	50963	41350	9		29289	2802	11402	3371	451	166	33	81610
1988	53058	44506	2		31457	2807	12568	3114	277	38	139	83.88
1989	54914	46458	8		32935	3504	12711	3613				84.61
1990	62131	55159	9	2	40114	9998	14450	7099	334	98	153	88.78
1991	60390	52359	9		51532	3684	10818	3233				87.40
1992	68338	5644	7		46961	4033	9010	3081	282	89	65	82.55
1993	69888	58878	7		48566	4493	9525	3841				84.25
1994	71667	61796	7		50921	4620	10875	4099				86.71
1995	88306	63689	5		51645	4587	11912	4059	69		58	87.1

(A) Year

(B) Number of married and fertile women

(C) Married and fertile women who use birth control

(C-1) Number of married and fertile women who use birth control
(C-2) controlled by male sterilization
(C-3) controlled by female sterilization
(C-4) controlled by IUD (Intra-Uterine Device)
(C-5) controlled by pills
(C-6) controlled by condom
(C-7) controlled by etc
(C-8) Birth Control Rate (%) $= \frac{(C\text{-}1)}{(B)}\%$
(C-9) All people who use birth control included
(C-10) Only those who newly started birth control this year included
(D) Unit: Person

Figure 2-C2: Birth Control Rate collected from local gazetteers

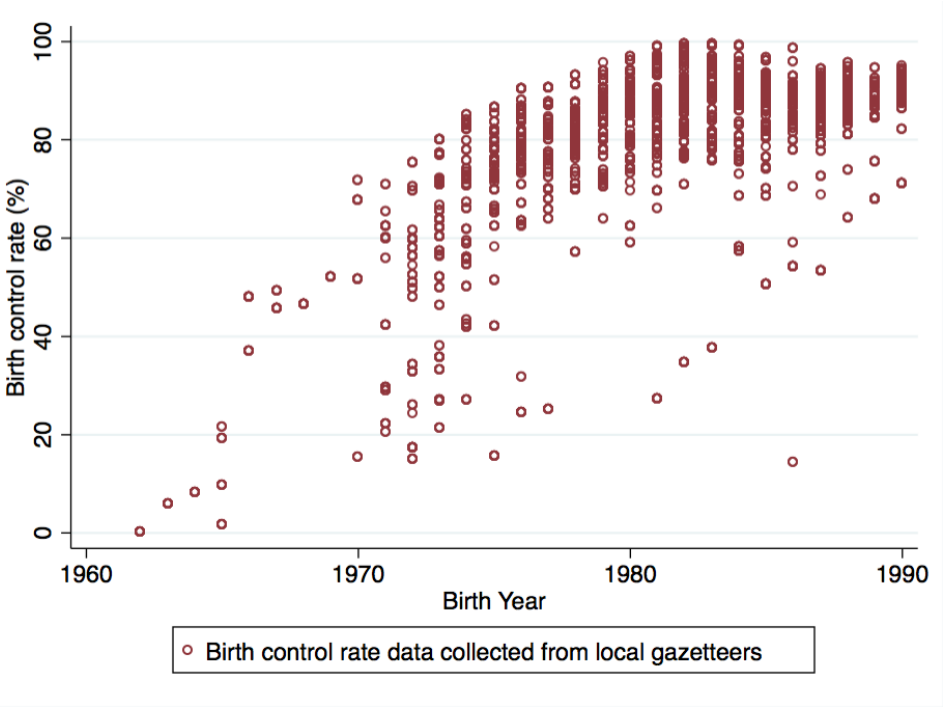


Figure 2-C2 plots the BCR we collected. In the early 1970s, it varies between about 20% to 80%, and the variation is smaller in later periods. Figure 2-C3 indicates that more than half of observations exceed 80% since 1977. That is, most of fertile women already were preventing conception before the legal introduction of One Child Policy in 1979.

Figure 2-C3: Birth Control Rate histogram by year (Unit: individual)

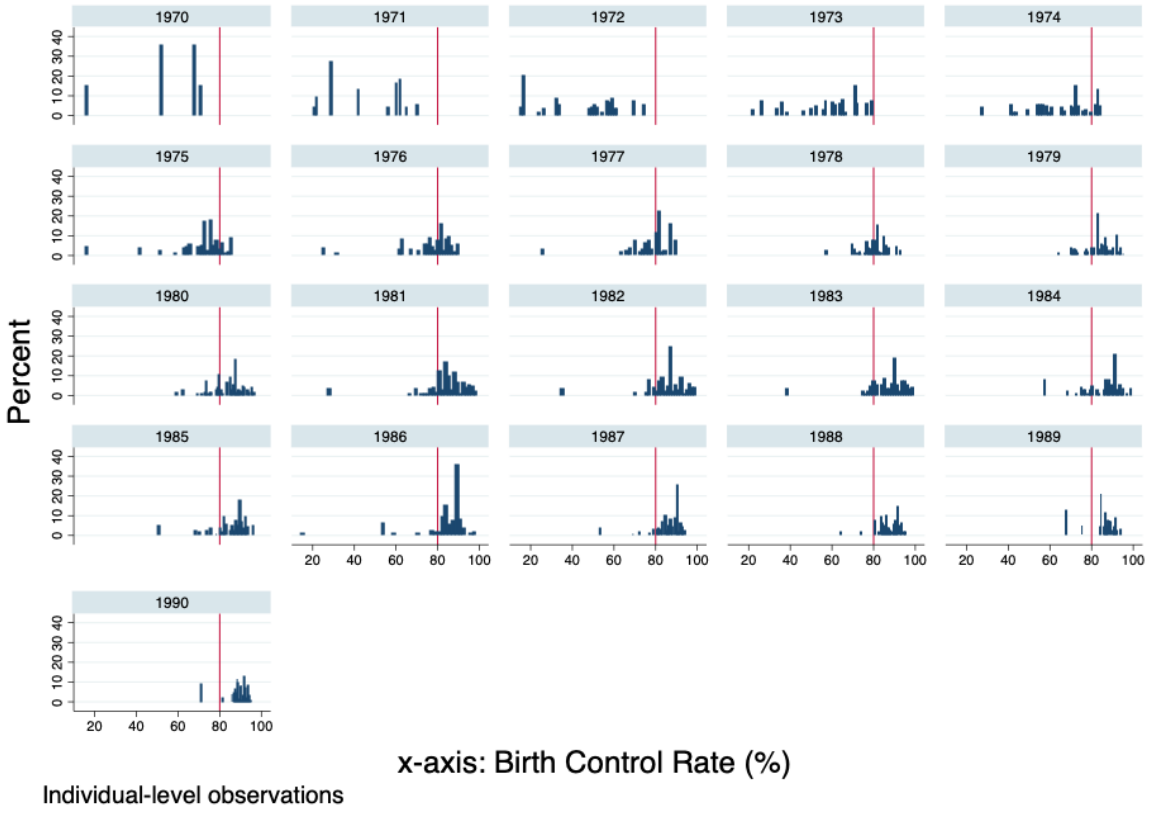


Table 2-C1: Birth Control Rate data availability by county

	Number of counties	%
County level info. available (A)	61	59.22%
Prefecture level info. available (B)	27	29.21%
Municipal urban/rural info. available (C)[1)]	15	14.56%
(A+B+C)	103	100%
Number of counties in the sample	162	

[1)] Shanghai only.

Table 2-C2 indicate that there are missing values, especially in earlier periods. The main reason is that the family planning policy was not intensive before 1970. Table 2-C2 (Panel A) indicates only 1.5% of rural individuals born in the 1960s have the BCR information of one's birth county in her birth year. On the other hand, more than 40% of individuals born between 1980 and 1990 and 22.6% of those born in the 1970s have the BCR information in the rural sample. Panel B in Table 2-C2 shows 63% of counties in the rural sample and 55% of counties in the urban sample have at least one year observation of the BCR. That is, although the number of available observations are limited, we still have at least one data point for more

than half of the counties, which is useful to predict the missing values.

Table 2-C2: Number of observations available in gazetteer data

	Rural Sample			Urban Sample		
Panel A						
Number of observations available						
Year	Available (A)	Total (B)	A/B (%)	Available (C)	Total (D)	C/D (%)
Total	3401	21641	15.7%	707	4058	17.4%
1980s	2089	4943	42.3%	515	1044	49.3%
1970s	1220	5402	22.6%	184	861	21.4%
1960s	92	6159	1.5%	8	1061	0.8%
1950s	0	5137	0.0%	0	1092	0.0%
Panel B						
Number of counties with at least an observation available						
Year	Available (E)	Total (F)	E/F (%)	Available (G)	Total (H)	G/H (%)
Total	102	162	63.0%	83	151	55.0%
1980s	100	162	61.7%	78	124	62.9%
1970s	69	162	42.6%	41	114	36.0%
1960s	4	159	2.5%	1	111	0.9%
1950s	0	162	0.0%	0	110	0.0%

Note:
1) 1980s covers from 1980 to 1990;

APPENDIX D: Survey-measured Birth Control Rate

The State Family Planning Commission in China conducted a deep survey of 459,269 ever-married women about their fertility and contraceptive histories: Two-per-thousand fertility and contraceptive survey (1988 Fertility survey). This survey asked each woman about their contraception experiences including contraception type, use year, and reasons why she used the contraception for each case. This data consists of 544,190 contraception case observations, and we use it to generate the Survey-measured BCR, which is $SurveyB_{pt}$ in equation (3). First, we aggregate contraception case observations by (1) prefecture and (2) contraception use year then compute the BCR as follows:

$$SurveyB_{pt} = \frac{\hat{CW}_{pt}}{\hat{W}_{pt}} * 100 \tag{2D.1}$$

where $SurveyB_{pt}$ is the Survey-measured BCR for prefecture p in year t. Simply, $SurveyB_{pt}$ is the proportion of women who use contraception in year t at prefecture p in the 1988 Fertility survey. $\hat{CW}_{pt}$ is the number of women who use contraception in year t, and $\hat{W}_{pt}$ is the total number of women who live in prefecture p in year t in the 1988 Fertility survey. $\hat{CW}_{pt}$ counts various contraceptive measures including (1) female sterilisation, (2) male sterilisation, (3) intrauterine contraceptive device (IUD), (4) oral contraceptive pills, (5) pills in external use, (6) injection, (7) condom, and (8) other medicine or medical devices to control births. Table 2-D3 indicates about 80% of those cases are either sterilisation or intrauterine device, which requires a surgery, rather than temporary one-off methods. A major difference between $SurveyB_{pt}$ and the actual BCR (B_{ct}) is the former does not count abortion cases. This leads $SurveyB_{pt}$ to be underestimated as the number of abortions also significantly grew under the family planning policy.

Table 2-D1 provides a simplified example of estimating the BCR using the 1988 Fertility survey. Since the unit of data is a case, most individuals have multiple observations. There are four main variables in the data - contraception type, start and end year, and the reason why one uses the contraception. In Table 2-D1, there are two women living in prefecture 1 so the denominator $\hat{W}_{1t} = 2$ in equation (2D.1) where $t = 1972, 1973, ..., 1988$. Women 1 used an intrauterine device between 1972 and 1973 so $\hat{CW}_{1,1972} = 1$ and $\hat{CW}_{1,1973} = 1$. That is, $\hat{W}_{pt}$ counts the number of women using contraceptive methods in year t at prefecture p. We only count cases that occurred to comply with the government family planning policy. For example, although Woman 1 uses contraceptive pills between 1975 and 1977, prefecture 1 has zero contraception cases between 1975 and 1997 as she used for personal (economic) reasons. Also, we do not count natural contraception cases when calculating $\hat{CW}_{pt}$ as it is not as effective as a tool- or surgery-based contraceptive methods. Table 2-D3 indicates the natural contraception accounts for 1.52% of 411,448 government contraception cases. Excluding natural conception cases, the final sample consists of 405,185 cases.

Table 2-D1 shows Woman 2 has a missing value in column 'End year'. We do not know

Table 2-D1: How to aggregate cases by prefecture and year

Prefecture	ID	Contraception type	Start year	End year	Reason
1	Woman 1	IUD	1972	1973	Government call
1	Woman 1	Pills	1975	1977	Economic reason
1	Woman 1	Injection	1979	1979	Government call
1	Woman 2	Natural contraception	1974	1974	Government call
1	Woman 2	Sterilisation	1981	.	Government call

⇓

Prefecture	Year	Number of women using contraception due to the government's policy or campaign	Number of women surveyed	
p	t	$C\hat{W}_{pt}$	$\hat{W}_{pt}$	$\hat{B}_{pt}(\%)$
1	1971	0	2	0
1	1972	1	2	50
1	1973	1	2	50
1	1974	0	2	0
1	1975	0	2	0
1	1976	0	2	0
1	1977	0	2	0
1	1978	0	2	0
1	1979	1	2	50
1	1980	0	2	0
1	1981	1	2	50
1	1982	1	2	50
...	...	...	...	...
1	1988	1	2	50

whether it means she was still using contraception or not responding. And Table 2-D4 shows that 58.9% (242,445 cases) of 411,448 cases have missing values on the year ended. We assume that missingness means the contraception did not end as of 1988. The reasons are as follows: (1) most contraception cases that started in the late 1980s do not have end year information. Figure 2-D1 plots the proportion of contraception cases that have end year information against its start year. Less than 30% of the cases that started after 1986 have end year observations. Overall the pattern in Figure 2-D1 indicates the earlier a case starts, the more likely it has the end year information. (2) Table 2-D4 shows that the proportion of missing values is significantly higher among relatively permanent contraception cases. 88.6% of female sterilisation and 79.4% of male sterilisation cases do not have the end year information, while the missing proportion of the rest are around 30-40%p lower. Another issue is that the survey was conducted in 1988 so we do not have observations for 1989 and 1990. We use 1988 observations for 1989 and 1990 by assuming that the BCR did not significantly change between those years.

Figure 2-D2 compares the Survey-measured BCR ($SurveyB_{pt}$) with the actual BCR from gazetteers. The Survey-measured BCR is lower than the actual BCR, possibly owing to the fact it does not count abortion cases. The correlation between the two variables is 0.6116.

Table 2-D2: Contraception case distribution by reason

Reason for contraception	Freq.	%
Economic reason	32,228	5.92
Family issue	46,016	8.46
Work-related	18,700	3.44
Education	1,223	0.22
Health	13,815	2.54
Government policy or campaign	411,448	75.61
Others	20,760	3.81
Total	544,190	100.00

Table 2-D3: Birth Control Rate estimation sample in the 1988 Fertility survey

	Freq.	%
Total number of contraception cases	544,190	
Contraception cases to comply with the government policy	411,448	100.00
- Male sterilisation	25,494	6.19
- Female sterilisation	102,827	25.00
- Intrauterine device	221,130	53.74
- Pills	35,744	8.69
- Condom	17,404	4.23
- Medical device or Medicine for external use	2,586	0.63
- Natural contraception	6,263	1.52
Estimation sample		
Number of cases excl. natural contraception	405,185	98.48

Table 2-D4: Missing values in contraception end year information

	End Year information		
	Available (A)	**Missing** (B)	(A+B)
Number of government contraception cases	169,003	242,445	411,448
%	41.08%	58.92%	100.00%
By contraception type			
Male sterilization	20.6%	79.4%	100.0%
Female sterilization	11.4%	88.6%	100.0%
IUD (Intra-Uterine Device)	53.8%	46.2%	100.0%
Pills	51.8%	48.2%	100.0%
Condom	53.5%	46.5%	100.0%
Medical device or Medicine for external use	61.6%	38.4%	100.0%
Natural contraception	49.5%	50.5%	100.0%

Figure 2-D1: The proportion of cases with end year observation

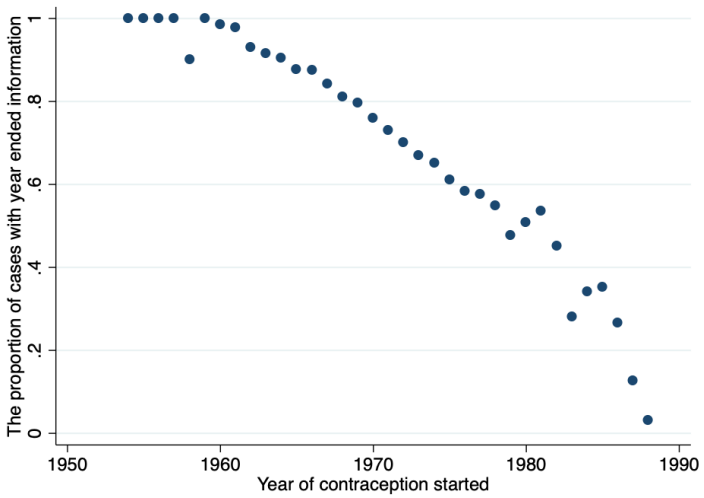
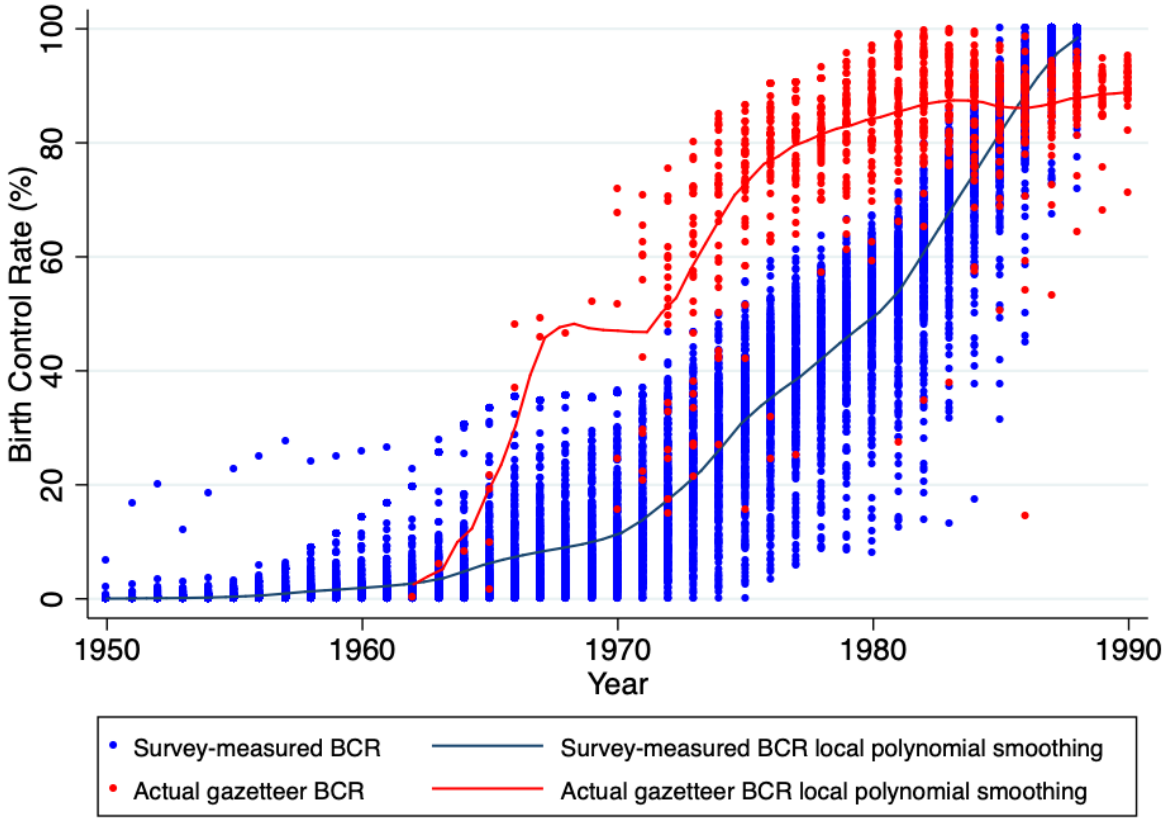


Figure 2-D2: Survey-measured BCR and the actual gazetteer BCR



APPENDIX E: Missing value Imputation

We estimate the BCR using the Survey-measured BCR ($SurveyB_{pt}$) as described in equation (3). Table 2-E1 provides the estimation results. Column 5 indicates that (1) the imputation model explains 78% of the variations in the actual gazetteer BCR and (2) 1% point increase in $SurveyB_{pt}$ is associated with 0.963% point increase in the actual BCR.

Table 2-E1: Estimation results used for imputation

Dependent var. : Actual gazetteer Birth Control Rate (%) B_{ct}					
	(1)	(2)	(3)	(4)	(5)
$SurveyB_{pt}$	0.277***	0.271***	0.929***	0.930***	0.963***
	[0.013]	[0.012]	[0.053]	[0.053]	[0.053]
Prefecture unit dummy				2.299	9.319***
				[2.688]	[2.955]
$SurveyB_{pt}$*Prefecture unit dummy					-0.096***
					[0.018]
Constant	64.883***	65.268***	30.650***	29.619***	26.688***
	[0.876]	[0.819]	[1.711]	[2.093]	[2.135]
County fixed effect	No	Yes	Yes	Yes	Yes
Year fixed effect	No	No	Yes	Yes	Yes
Observations	1,077	1,077	1,077	1,077	1,077
R-squared	0.308	0.573	0.774	0.774	0.781

Note:
1) Robust Standard Errors (SEs) are presented in parentheses;
2) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;
3) Test year and county information source indicators are included in column (3)-(5);
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

Figure 2-E1 and 2-E2 compare the actual BCR observations collected from local gazetteers with imputed values according to equation (3). Both Figures suggest that imputed values are similar to the original values.

Figure 2-E1: Actual values vs imputed values of Birth Control Rate

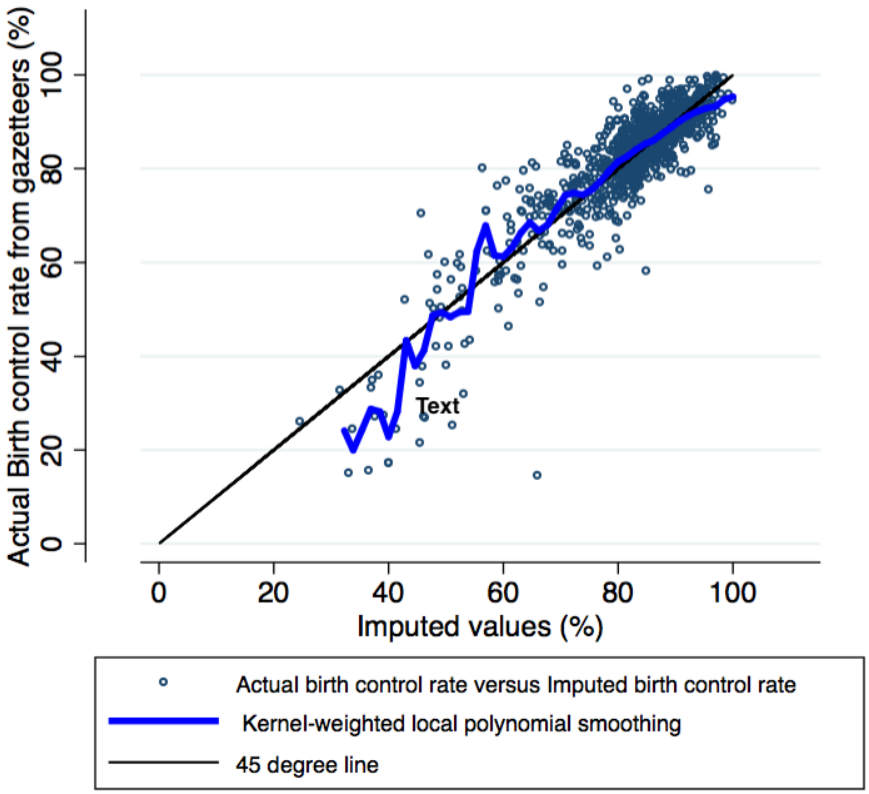


Figure 2-E2: Actual values vs Imputed values of Birth Control Rate

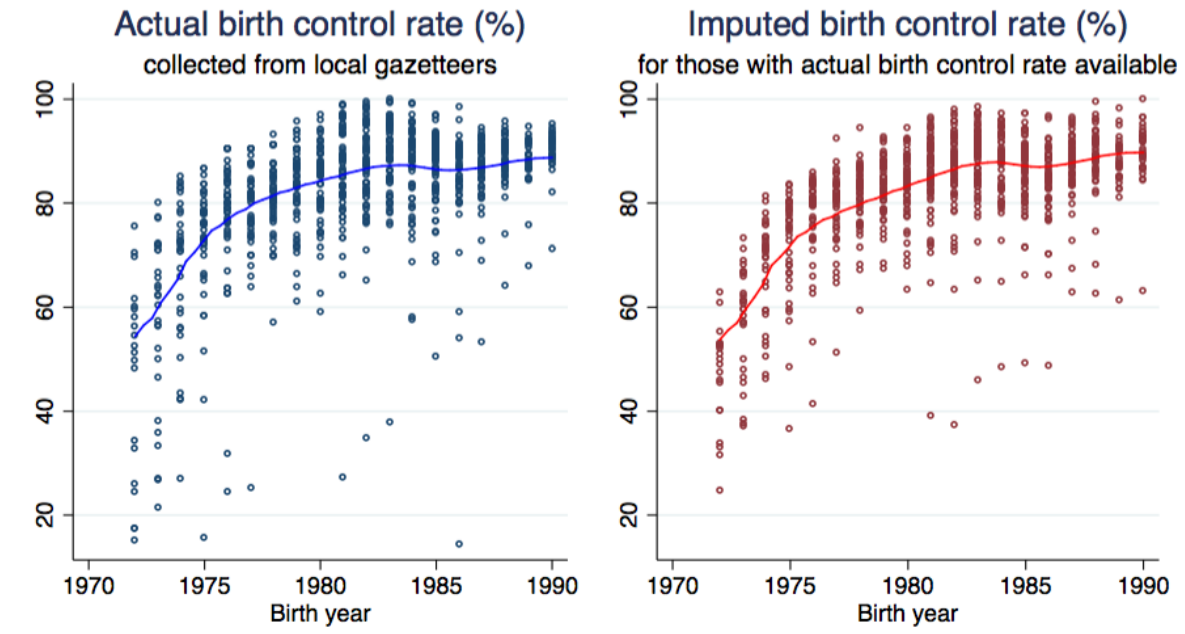


Figure 2-E3: Birth Control Rate at one's birth

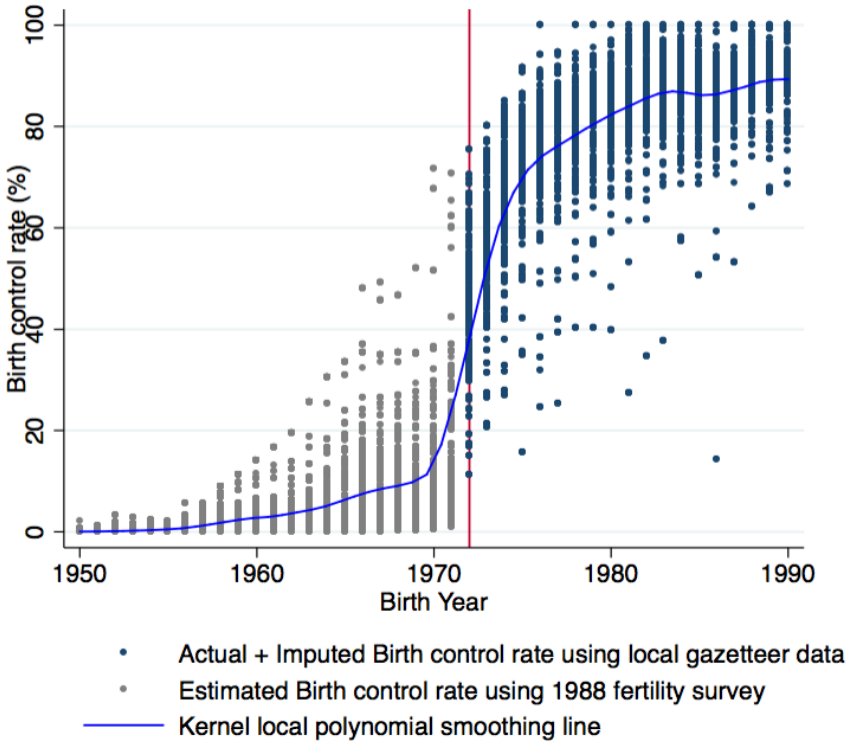
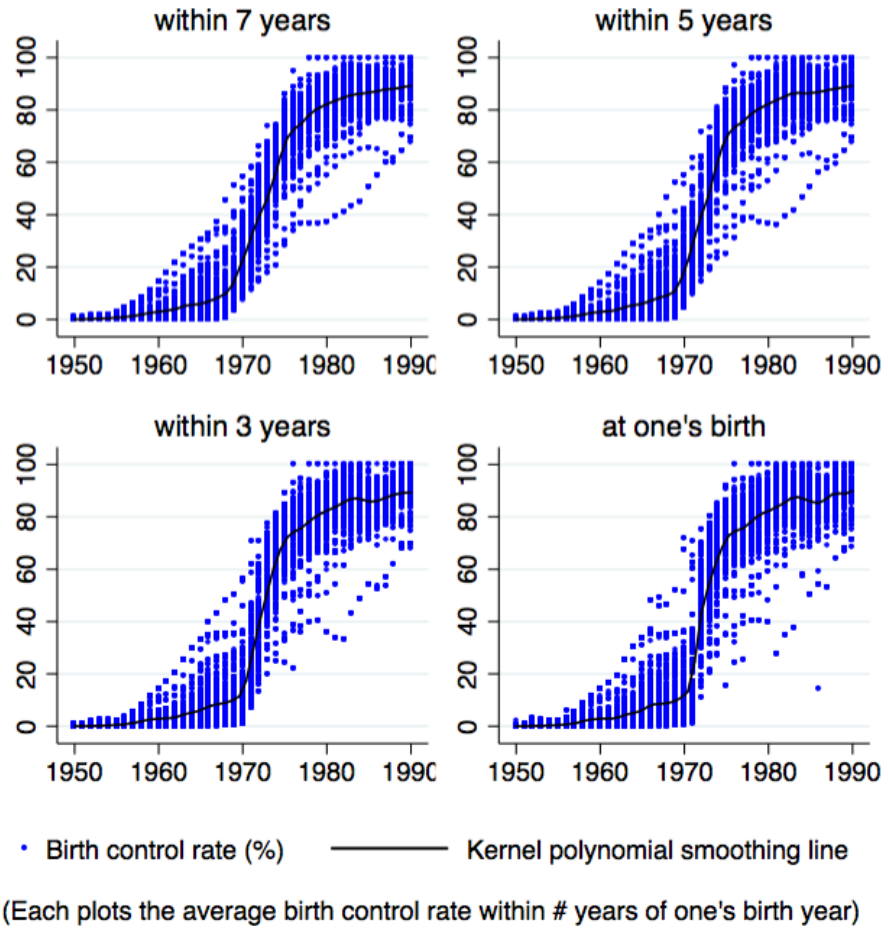


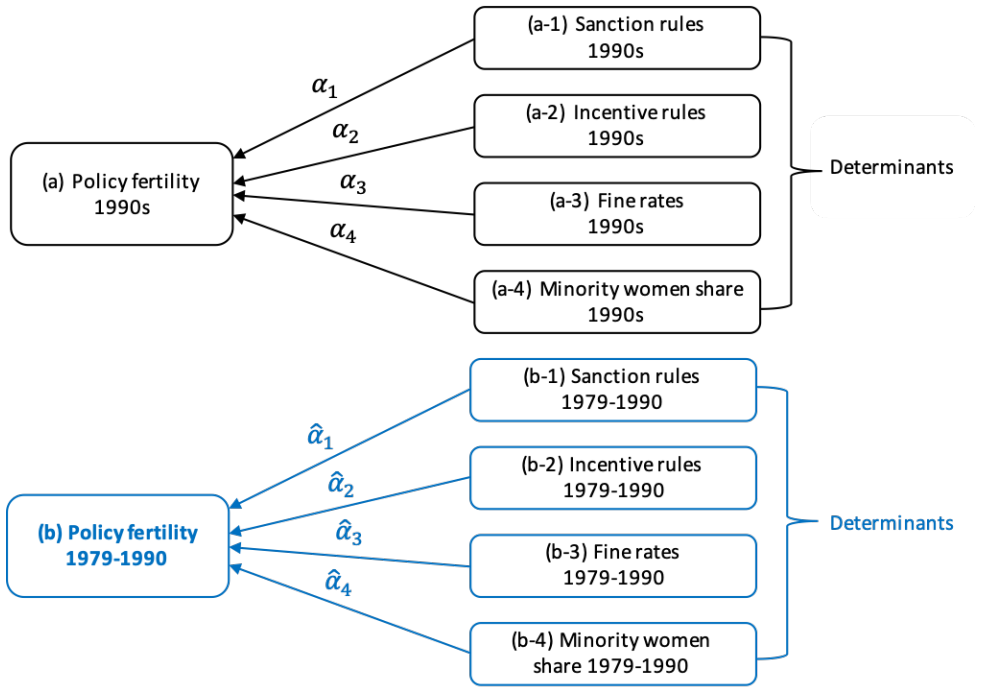
Figure 2-E4: Average Birth Control Rate within 1/3/5/7 years



(Each plots the average birth control rate within # years of one's birth year)

APPENDIX F: One Child Policy target fertility index

Figure 2-F1: How to measure the OCP target fertility at the prefecture level



The second instrument is a new index which measures the government's fertility target (Policy Fertility Index; PFI) at the prefecture level under the OCP on an annual basis. To do so, we collect several data of the OCP details from literature and predict the fertility level each local government targeted in the past.

Figure 2-F1 summaries how to estimate the policy fertility between 1979 and 1990. The goal is to predict '(b) Policy fertility 1979-1990', and we use data on (a), (a-1), (a-2), (a-3), (a-4), (b-1), (b-2), (b-3), and (b-4). First, we regress (a) on (a-1), (a-2), (a-3), and (a-4) to estimate α_1, α_2, α_3 and α_4 as stated in equation 4. Second, we apply the estimated αs (i.e. $\hat{\alpha}$s) to the four determinants (b-1), (b-2), (b-3) and (b-4) and finally (b). This method requires an assumption that each prefecture updates their family planning policies in agreement with its provincial government policy updates as we assume provincial-level incentive, sanction and fine variables are highly associated with prefecture-level policy fertility. Given that cities and prefectures follow new scheme and directions made by the provincial government [Scharping, 2003], this assumption is reasonable. Data sources are available in Table 2-F1.

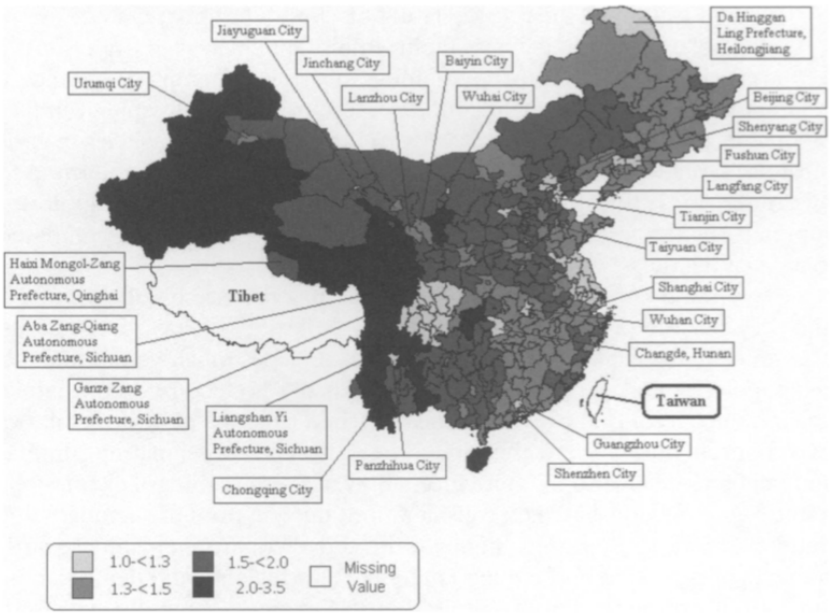
Figure 2-F1 (a): Gu et al. [2007] provides the local government's fertility target information at the prefecture level as of the 1990s. They surveyed each local government and constructed a data set of the fertility control policy details for 420 prefectures as of the 1990s. Then they merge this data with the 2000 Population Census and estimate the number of children that each couple was allowed to have in the 1990s, which summarises each prefecture's birth control policy intensity. This variable is called 'policy fertility'. Figure 2-F2 shows its geographical distribution. Figure 2-F3 plots its distribution among our 162 sample counties. About 20% of

Table 2-F1: Data used to predict the Policy Fertility Index

Source	Data	Variable	Unit	Available
Gu et al. (2007)	Policy fertility	P	Prefecture	A single year (=1990s)
Scharping (2003)	Sanction/Incentive rules	S, I	Province	Every year in 1979-1999
Ebenstein (2010)	Fine rates in PDV	F	Province	Every year in 1979-1999
China Census (1982/1990)[1)]	Minority share among 15-45 aged women	M	prefecture	Every year in 1979-1990

the counties have 1.11 policy fertility levels in the 1990s. That is, a couple in those counties was allowed to have up to 1.11 children on average. More than 50% of the 162 counties have the policy fertility greater than 1.14 and less than 1.5 children. About 27% lies between 1.5 and 2, and the rest (1.24%) exceeds 3.

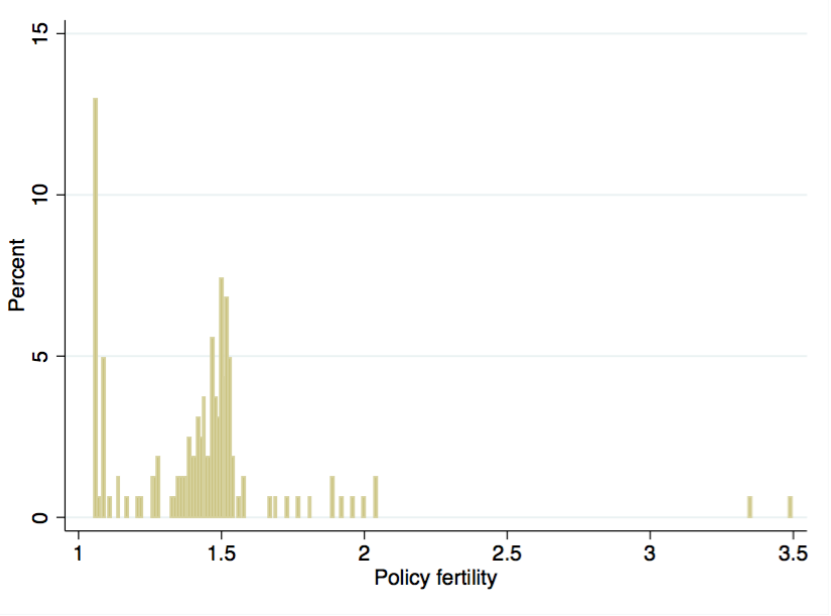
Figure 2-F2: Geographical distribution of the policy fertility



Source: Gu et al. (2007) Figure 2

Figure 2-F1 (a-1), (a-2): As described in equation (4) in the main text, we collected provincial-level data on four variables that are highly correlated with the policy fertility. The first two variables (a-1) and (a-2) are sanction rules on unauthorised births and incentive rules for one-child families, collected from Scharping [2003]. Figure 2-F2 shows a part of the tables that we collected data from. 'Date' stands for the date of policy updates and the numbers in columns 'Urban' and 'Rural' indicate types of polices introduced on the date. Table 2-F2 provides the list of policy types provided in Scharping [2003]. We generate policy dummy variables which

Figure 2-F3: Policy fertility distribution in our sample counties



indicate whether a policy implemented in a year at a province. Considering that some rules could be effective only either in urban or rural areas, we include different sets of variables in the urban and rural analyses. For example, an incentive rule is to allocate more allotment of farm land to one-child families. This policy should be effective in rural regions but not in urban areas. Table 2-F2 highlights incentive and sanction rule indicators that are included in the prediction model. Some rule indicators do not vary across time or provinces because those rules either never implemented or always implemented. Table 2-F3 shows the average of each rule indicator in the 1990s. The sanction rules 16, 17, and 18 have zero mean in urban areas, meaning no province had these rules in the 1990s. All provinces had the incentive rule 1 in the 1990s.

Figure 2-F4: Province birth control policy information in Scharping (2003)

Table 10 Provincial birth-planning rules: one-child incentives 1979–99

	Date	*Urban*	*Rural*
Beijing	6-12-82	1 (60 ¥),2-5,7	1 (60 ¥) or 12 or 14;9
	15-1-91	1 or 2;3-6	1,3,7,15
Tianjin	14-10-82		
	2-11-88	1,3-5	1 or 7,15;3
	9-3-93		
	15-4-94		
	30-1-97	1,3-5	1 or 7;15;13
Hebei	5-4-82	1 (24-30 ¥),2-5,7	1 (15-20 ¥) or 12,14;3-5
	28-2-86		
	14-3-89	1 (60 ¥),2-5	1 (60 ¥),3-5,7,13

Source - Scharping (2003) Table 10

Figure 2-F1 (a-3): The third determinant is a fine rate that are applied to unauthorised births. Ebenstein [2010] uses data from Scharping [2003] and estimate fine rates by province and year. Specifically, this variable is the number of years of annual income that one needs to pay for fines to have an excess fertility. For example, if the fine rate is 2 in this data, the amount of fine that one needs to pay is equivalent to 2 years of one's annual income. See page 111 in Ebenstein [2010] for more details.

Figure 2-F1 (a-4): The forth variable is the proportion of minority women (non-Han Chinese) among fertile women (between age 15 and 45). We include this variable because ethnic minorities are allowed to have more children in general, so the proportion of ethnic minorities are highly correlated with policy intensity in the region [Cameron and Meng, 2014]. To construct this variable, we use (data 1) China 1% Population Census 1982, (data 2) China 1% Population Census 1990 and (data 3) China One-Per-Thousand Census 2000. First, we count the number of women aged between 15 and 45 (fertile age) by year and prefecture. For example, this number in 1985 is equal to the number of women born between 1940 and 1970 in (data 2) as their ages are between 15 and 45 in 1985. Second, we count the number of ethnic minority women aged between 15 and 45 by year and prefecture. Since (a-4) in Figure 2-F1 is about the 1990s, we use (data 3) to generate this variable. Figure 2-F6 shows distributions of this variable.

Table 2-F2: Incentive and Sanction rule information in Scharping (2003)

Incentives for one-child families:
1 one-child bonus or health insurance fee up to age 14-18
2 extension of maternity leave or better financial conditions for paid maternity leave
3 preference in the allocation of housing or the allotment of building plots (u) (r)
4 preference in placement for nurseries and kindergartens and in school enrolment (u) (r)
5 preference in medical care and hospitalisation (r)
6 paid nursing or caretaking leave (u)
7 preference in recruitment for non-agricultural jobs (r)
8 preference in army recruitment
9 extra pension for employees or assistance for old peasants (u) (r)
10 pensioners with only one daughter covered by medical insurance of daughter or son-in-law
11 higher subsidised food ration
12 more allotment of farm land (r)
13 preference in credit, relief and economic assistance (r)
14 reduced procurement quotas of low-priced grain and other products, less collective fees
15 exemption from or reduction of compulsory labour
16 reward for areas or work units with commendation for birth planning
17 old-age insurance for peasants (r)
Sanctions for unauthorised births:
1 fines for illegitimate or early births, neglect of spacing rules or disregard of formalities for & permitted first, second or higher-order births (r)
2 above-quota birth fines (regular or one-time income deductions for both parents of non-authorised child) (r)
3 social-support fee for non-authorised birth, one-time or in installment
4 more severe wage and income deductions or fines for third and higher-order child (u) (r)
5 privately paid medical and maternity expenses, no paid maternity leave (u)
6 no bonus payments, welfare items or commendation benefits (limited in time)
7 no job promotion (limited in time)
8 demotion
9 possible revocation of business license for persons self-employed or holding provisional & registration (r)
10 possible abrogation of leases of state or collective enterprises
11 possible discharge of non-tenured employees (u)
12 no medical insurance and social welfare benefits for unauthorised children (limited in time), no relief (u)
13 no subsidised grain rations
14 no higher allocation of housing or allotment of building tracts (u) (r)
15 no higher allotment of farm land, no social relief (r)
16 reduction of land allotments
17 higher procurement quotas for low-priced grain and other produce
18 higher fees for the collective
19 no work allocation in state or collective enterprises for peasants or urban unemployed (r)
20 no non-agricultural household registration or provisional registration in urban areas (r)
21 disciplinary sanctions for cadres and state employees up to discharge from work
22 public criticism and account of cadres
23 fines for responsible cadres
24 economic sanctions and no commendation benefits for work unit or area (u) (r)
25 further sanctions according to birth-planning contract (u)

Note:
1) Source - Scharping (2003) Table 10 and 11;
2) (u) means this policy indicator is included in the prediction model for urban sample;
3) (r) means this policy indicator is included in the prediction model for rural sample;

Table 2-F3: Average of Incentive and Sanction policy indicators in 1990s

Sanction rule No.	Urban	Rural	Incentive rule No.	Urban	Rural
1	0.848	0.729	1	1.000	0.901
2	0.994	0.923	2	0.286	0.000
3	0.076	0.056	3	0.906	0.878
4	0.894	0.865	4	0.809	0.488
5	0.648	0.156	5	0.785	0.415
6	0.687	0.206	6	0.006	0.000
7	0.624	0.000	7	0.515	0.817
8	0.236	0.000	8	0.000	0.000
9	0.043	0.109	9	0.498	0.369
10	0.099	0.114	10	0.020	0.000
11	0.006	0.006	11	0.031	0.019
12	0.288	0.089	12	0.000	0.222
13	0.018	0.001	13	0.000	0.428
14	0.420	0.564	14	0.000	0.044
15	0.006	0.180	15	0.000	0.123
16	0.000	0.012	16	0.133	0.025
17	0.000	0.000	17	0.000	0.041
18	0.000	0.000			
19	0.148	0.309			
20	0.000	0.307			
21	0.836	0.118			
22	0.105	0.099			
23	0.278	0.090			
24	0.530	0.367			
25	0.025	0.000			

Note:
1) See Table 2-F2 for policy description

Figure 2-F5: Fine rate distribution from Ebenstein (2010)

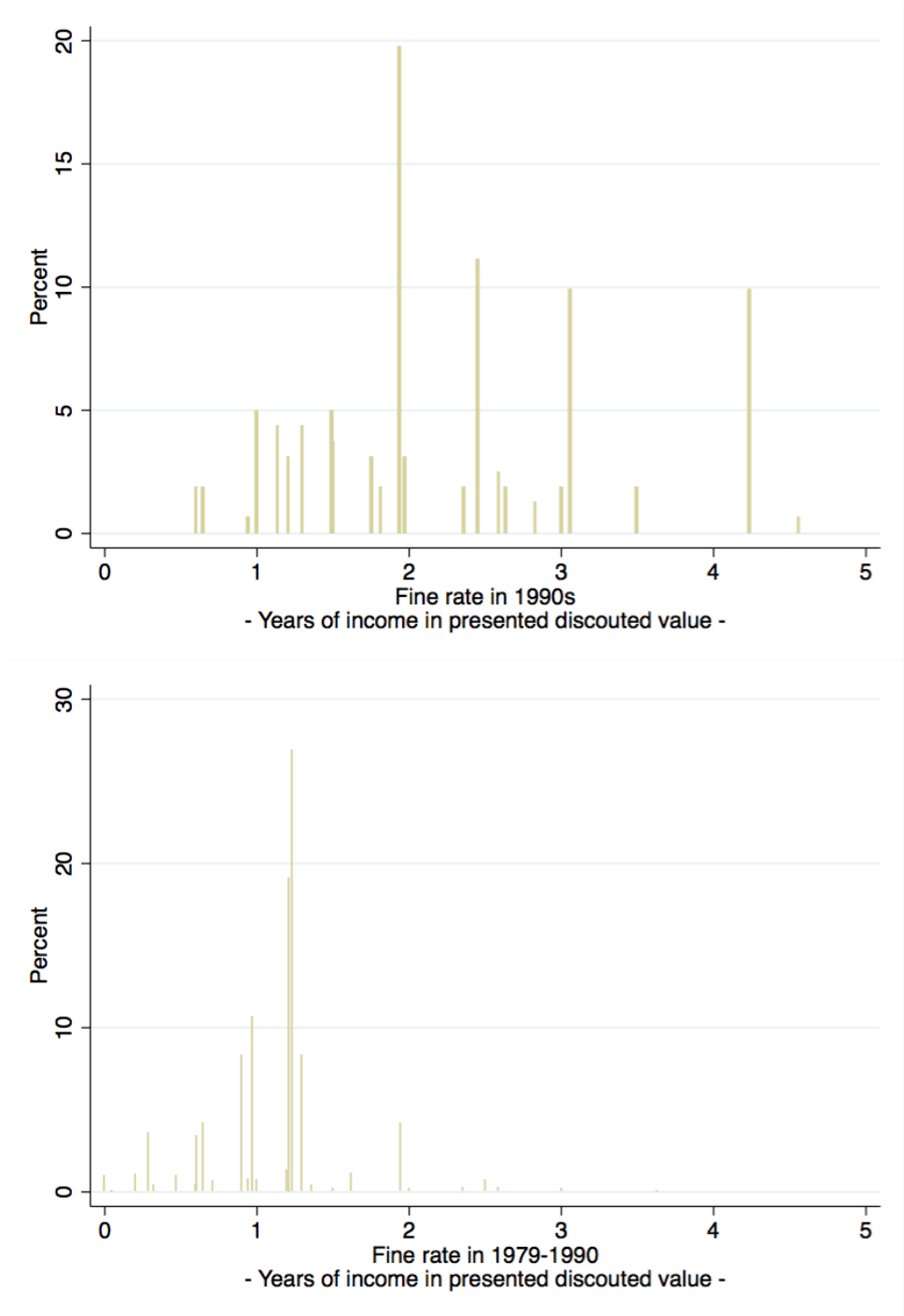


Figure 2-F6: Ethnic minority share of women aged between 15 and 45

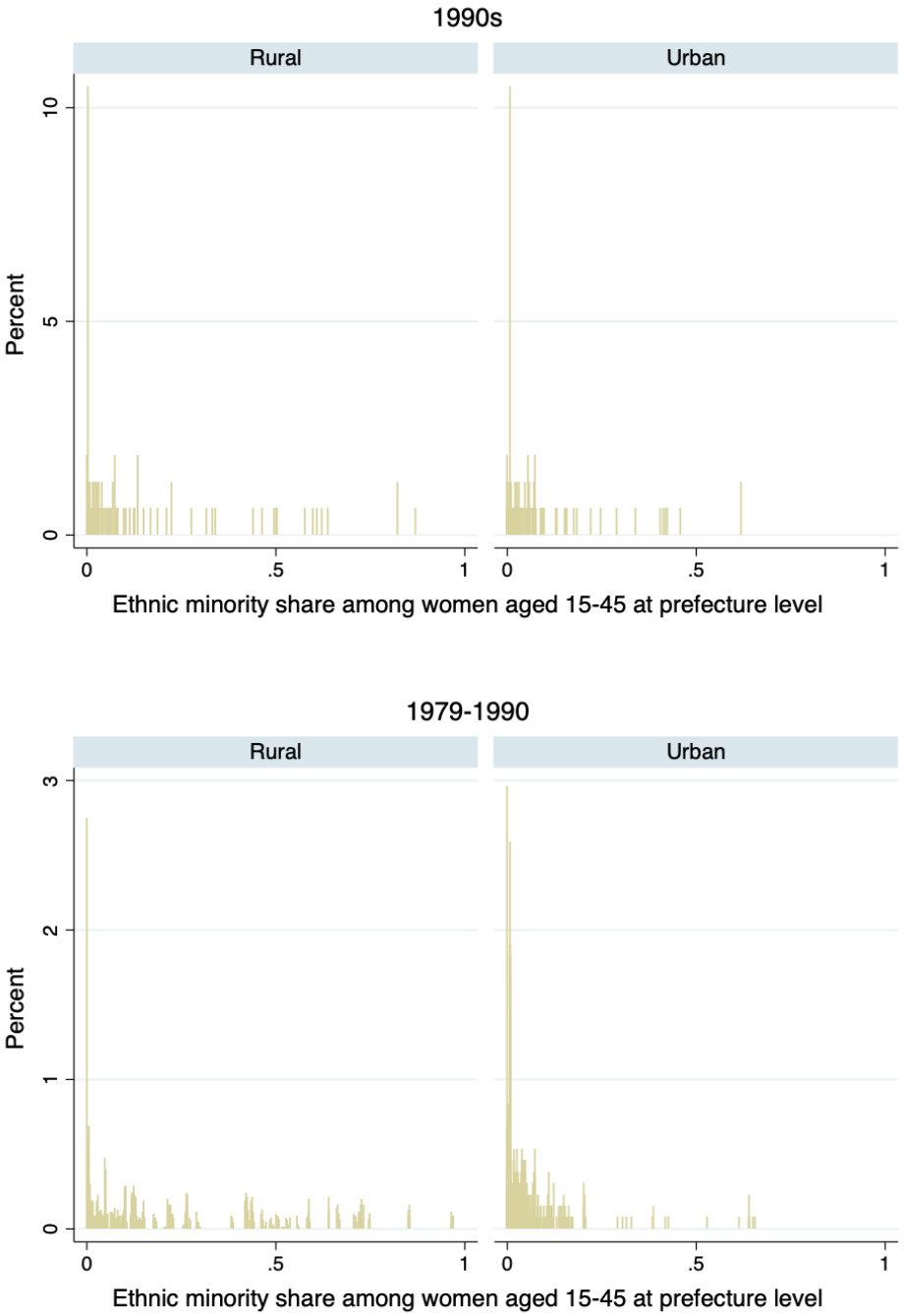


Figure 2-F1 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$: αs in equation (4) capture to what extent each determinant affects the outcome variable policy fertility. We estimate the αs by regressing (a) on (a-1), (a-2), (a-3), and (a-4) as mentioned earlier in equation (4). Table 2-F3 provides the estimation results. Column (2) shows fine rate is negatively associated with the number of children that a couple is allowed to have both in urban and rural regions in the 1990s. The higher ethnic minority share is, the more kids people are allowed to have. This estimation model explains the policy fertility variations by 29.17% for urban and 34.98% for rural regions in the 1990s.

Figure 2-F1 (b-1), (b-2), (b-3), (b-4): We have annual data on the four determinants for the period between 1979 and 1990 from the same data sources for (a-1), (a-2), (a-3), and (a-4).

Figure 2-F1 (b): To predict (b), we use the following equation:

$$\hat{P}_{t,f} = \hat{\alpha}_0 + \sum_{k=1}^{K} \hat{\alpha}_{1,k} S_{k,t,p,u} + \sum_{l=1}^{L} \hat{\alpha}_{2,l} I_{l,t,p,u} + \hat{\alpha}_3 F_{t,p,u} + \hat{\alpha}_4 M_{t,f,u} + \nu_{t,f,u} \tag{2D.2}$$

$$t = 1979, 1980, ..., 1990$$

where $\hat{P}$ represents the predicted number of children that a couple is allowed to have under the OCP in prefecture f in year t. S_k is the kth sanction rule indicator among K indicators, $I_{l,t,p}$ is the lth incentive rule indicator among L indicators in province p in year t, and F indicates the fine rate for an unauthorised birth in province p in year t. M is the share of ethnic minorities among fertile women in year t for prefecture f.

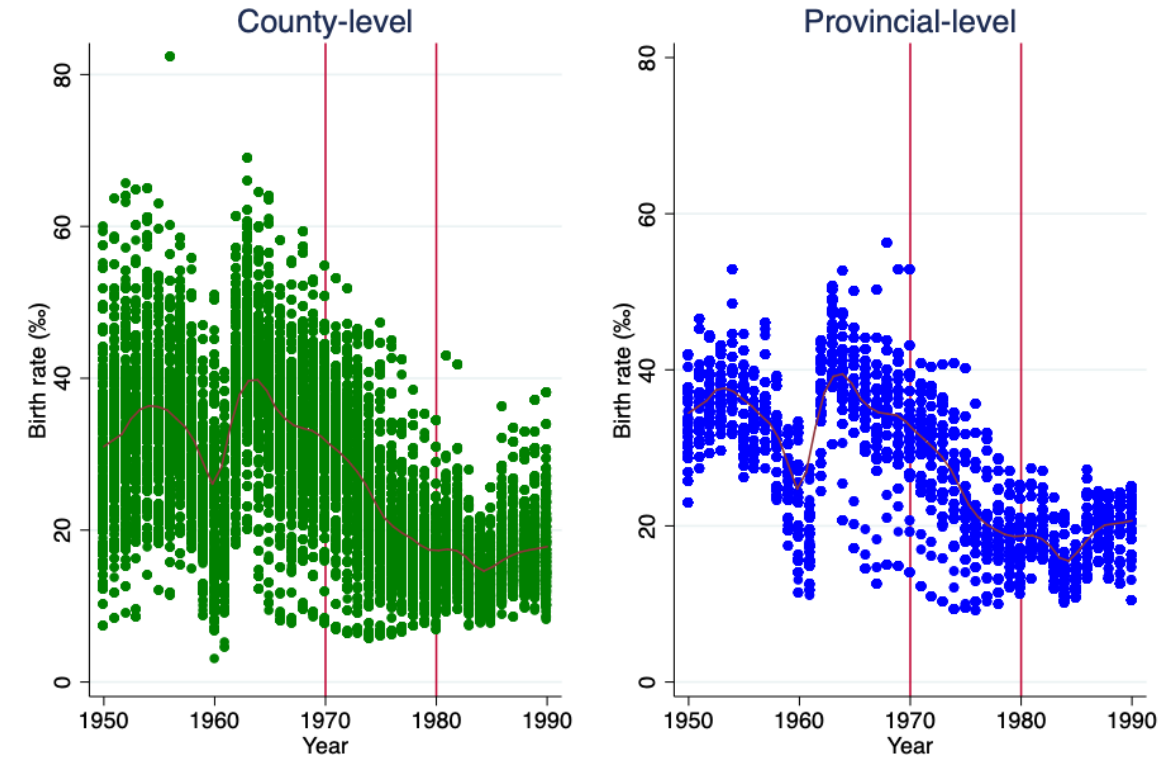
Since the predicted policy fertility only covers the period after the OCP (1979-1990), we do not have any observations of the policy fertility for the period before the OCP. We instead use county-level birth rates for the period between 1950 and 1978, collected from county gazetteers and provincial-level population statistical yearbooks (See Figure 2-F7). Therefore, the final instrument 'the Policy Fertility Index' consists of two parts: the predicted policy fertility for 1979-1990 and county-level birth rates for 1950-1978.

Table 2-F4: Estimating α

	Urban				Rural		
	(1) Selected	(2) Selected	(3) All		(1) Selected	(2) Selected	(3) All
Sanction 4	0.048	0.260**	0.294	Sanction 1	-0.033	0.108	-0.977**
	[0.125]	[0.104]	[0.187]		[0.109]	[0.079]	[0.435]
Sanction 5	0.051	0.080	0.245	Sanction 2	0.008	0.541**	
	[0.074]	[0.060]	[0.226]		[0.369]	[0.269]	
Sanction 11	-0.201	0.032		Sanction 4	0.539***	0.173	2.453***
	[0.380]	[0.307]			[0.164]	[0.122]	[0.921]
Sanction 12	-0.170*	-0.284***	-0.161	Sanction 9	-0.042	0.536***	3.051***
	[0.092]	[0.075]	[0.309]		[0.205]	[0.156]	[0.978]
Sanction 14	0.029	-0.003	0.200	Sanction 14	-0.317*	0.068	-0.573**
	[0.120]	[0.097]	[0.181]		[0.185]	[0.137]	[0.267]
Sanction 24	-0.202*	-0.031	0.098	Sanction 15	-0.210**	-0.031	-1.526**
	[0.117]	[0.096]	[0.163]		[0.096]	[0.071]	[0.623]
Sanction 25	-0.310*	-0.286**	-0.837***	Sanction 19	-0.294	0.24	
	[0.165]	[0.132]	[0.254]		[0.317]	[0.232]	
				Sanction 20	0.264	-0.038	-0.500*
					[0.234]	[0.170]	[0.278]
				Sanction 24	-0.148*	-0.101*	-1.964**
					[0.085]	[0.061]	[0.777]
Incentive 3	0.192	0.115	-0.019	Incentive 3	0.049	-0.033	-1.087**
	[0.165]	[0.133]	[0.341]		[0.132]	[0.095]	[0.426]
Incentive 4	-0.227*	-0.140	-0.068	Incentive 4	0.064	0.064	1.202***
	[0.123]	[0.099]	[0.174]		[0.119]	[0.085]	[0.438]
Incentive 9	-0.047	0.085	-0.075	Incentive 5	-0.078	0.213**	-0.451
	[0.072]	[0.060]	[0.183]		[0.145]	[0.107]	[0.324]
				Incentive 7	-0.217	-0.087	-2.257***
					[0.354]	[0.255]	[0.863]
				Incentive 9	-0.251***	0.118	0.251**
					[0.089]	[0.071]	[0.102]
				Incentive 12	0.063	0.199*	1.614***
					[0.160]	[0.116]	[0.543]
				Incentive 13	0.421***	0.13	2.577***
					[0.145]	[0.107]	[0.962]
				Incentive 17	0.141	-0.305	-1.866***
					[0.271]	[0.198]	[0.660]
Fine rate	0.000	-0.072***	0.000	Fine rate	0.032	-0.093**	0.221
	[0.033]	[0.028]	[0.070]		[0.055]	[0.041]	[0.144]
Minority share		1.793***	2.100***	Minority share		1.506***	1.515***
		[0.197]	[0.259]			[0.130]	[0.138]
Constant	1.538***	1.275***	0.631**	Constant	1.219**	0.513	1.691***
	[0.163]	[0.134]	[0.292]		[0.530]	[0.385]	[0.304]
Observations	162	162	162	Observations	162	162	162
R-squared	0.173	0.468	0.580	R-squared	0.292	0.637	0.669

Note:
1) Robust Standard Errors (SEs) are presented in parentheses;
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;
3) Columns (1) and (2) include only selected sanction and incentive rule indicators presented in this table;
4) Column (3) includes all available sanction and incentive rule indicators presented in Table 2-F2
5) See Table 2-F2 for policy description;

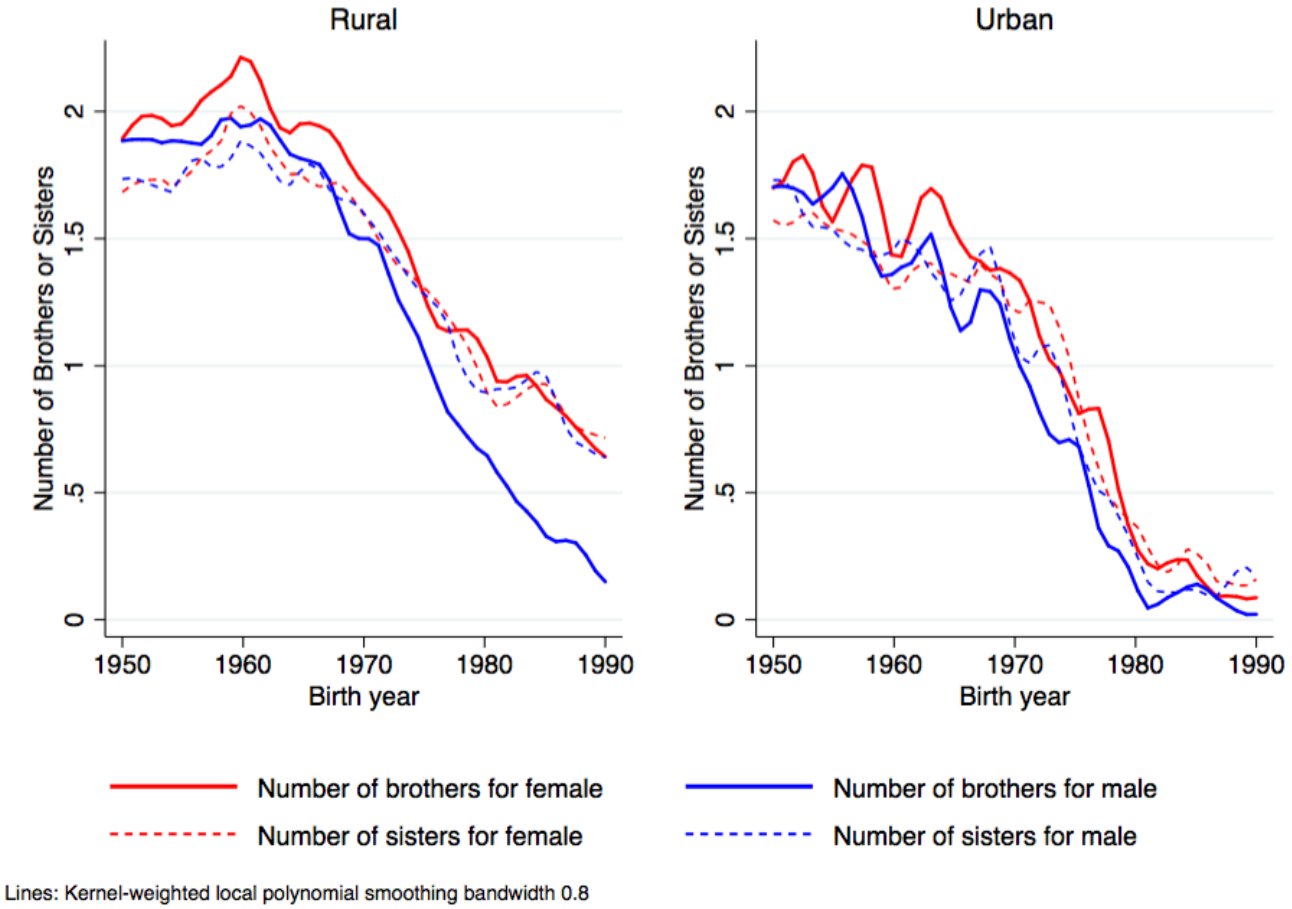
Figure 2-F7: County-level birth rates



Red line: Kernel-weighted local polynomial smoothing with Bandwidth 1

APPENDIX G: Figures and Tables

Figure 2-G1: Number of brothers and sisters



Lines: Kernel-weighted local polynomial smoothing bandwidth 0.8

Source: China Family Panel Survey 2010 and 2014

Table 2-G1: OLS results - Rural sample (continued)

Dependent var.	Panel A: Rural sample									
Vocabulary test z-score	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Siblings	-0.124***	-0.124***	-0.113***	-0.014***	-0.007*	-0.010**	-0.007	0.000	0.007**	0.000
	[0.006]	[0.006]	[0.005]	[0.004]	[0.004]	[0.005]	[0.005]	[0.003]	[0.003]	[0.004]
Siblings*Female										
Female	-0.311***	-0.311***	-0.322***	-0.342***	-0.349***	-0.280***	-0.283***	-0.081***	-0.080***	-0.389***
	[0.022]	[0.022]	[0.021]	[0.021]	[0.021]	[0.026]	[0.026]	[0.015]	[0.015]	[0.022]
Age				-0.017***	-0.005	-0.005	-0.001	0.007**		0.000
				[0.005]	[0.005]	[0.005]	[0.005]	[0.003]		[0.005]
Age squared				-0.000***	-0.000***	-0.000***	-0.000***	-0.000***		-0.000***
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]		[0.000]
Han Chinese					0.122*	0.122*	0.120*	0.075**	0.076**	0.120*
					[0.063]	[0.063]	[0.064]	[0.034]	[0.033]	[0.064]
Parental schooling years					0.049***	0.049***	0.050***	0.006***	0.004***	0.050***
					[0.003]	[0.003]	[0.002]	[0.002]	[0.002]	[0.002]
Birth order						0.019***	0.017***	0.004	-0.002	
						[0.006]	[0.005]	[0.004]	[0.004]	
Birth order*Female						-0.027***	-0.027***	-0.010**	-0.009*	
						[0.007]	[0.007]	[0.005]	[0.005]	
Individual schooling years								0.134***	0.137***	
								[0.002]	[0.002]	
Firstborn										-0.009
										[0.015]
Firstborn*Female										0.109***
										[0.020]
Birth county dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth county dummies*time	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Birth year dummies	No	No	No	No	No	No	No	No	Yes	No
Observations	21,641	21,641	21,641	21,641	21,641	21,641	21,641	21,641	21,641	21,641
R-squared	0.100	0.100	0.228	0.379	0.400	0.400	0.414	0.687	0.693	0.415

Note:
1) The outcome variables are test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample

Table G1: OLS results - Urban sample (continued)

Dependent var.	Panel B: Urban sample									
Vocabulary test z-score	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Siblings	-0.098***	-0.098***	-0.087***	-0.016*	-0.002	-0.001	0.003	0.014*	0.012	0.002
	[0.007]	[0.007]	[0.007]	[0.009]	[0.009]	[0.010]	[0.010]	[0.008]	[0.008]	[0.009]
Siblings*Female										
Female	0.006	0.006	-0.005	-0.010	-0.020	-0.006	-0.011	-0.023	-0.020	-0.025
	[0.015]	[0.015]	[0.014]	[0.014]	[0.015]	[0.026]	[0.026]	[0.025]	[0.023]	[0.022]
Age				0.004	0.004	0.004	0.446***	0.317***		0.446***
				[0.005]	[0.005]	[0.005]	[0.017]	[0.014]		[0.017]
Age squared				-0.000***	-0.000***	-0.000***	-0.000***	0		-0.000***
				[0.000]	[0.000]	[0.000]	[0.000]	[0.000]		[0.000]
Han Chinese					0.066	0.067	0.059	0.009	0.02	0.059
					[0.063]	[0.063]	[0.065]	[0.043]	[0.044]	[0.064]
Parental schooling years					0.035***	0.035***	0.035***	0.008***	0.006**	0.035***
					[0.003]	[0.003]	[0.003]	[0.002]	[0.002]	[0.003]
Birth order						0.003	-0.003	-0.014	-0.015	
						[0.012]	[0.012]	[0.011]	[0.010]	
Birth order*Female						-0.006	-0.004	0.011	0.01	
						[0.010]	[0.010]	[0.009]	[0.008]	
Individual schooling years								0.104***	0.105***	
								[0.003]	[0.003]	
Firstborn										0.005
										[0.028]
Firstborn*Female										0.014
										[0.032]
Birth county dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth county dummies*time	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Birth year dummies	No	No	No	No	No	No	No	No	Yes	No
Observations	4,058	4,058	4,058	4,058	4,058	4,058	4,058	4,058	4,058	4,058
R-squared	0.091	0.091	0.183	0.259	0.297	0.297	0.333	0.543	0.555	0.333

Note:
1) The outcome variables are test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample.

Table 2-G2: Years of schooling as the outcome variable

	Rural sample		Urban sample	
Dep. var.	(OLS)	(IV)	(OLS)	(IV)
Years of schooling	(1)	(2)	(3)	(4)
Siblings	0.167***	1.331***	0.019	0.794
	[0.031]	[0.191]	[0.073]	[0.952]
Siblings*Female	-0.431***	-1.259***	-0.226***	-0.317**
	[0.040]	[0.087]	[0.085]	[0.131]
Female	-1.080***	-0.439***	0.124	0.067
	[0.128]	[0.154]	[0.162]	[0.181]
Controls	Yes	Yes	Yes	Yes
Observations	21,641	21,641	4,058	4,058

Note:
1) The outcome variable is individual schooling years when a respondent was surveyed;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in the main IV estimation in Table 4;
4) The instrument is the Birth control rate;

Table 2-G3: Sibling numbers by gender and relative birth order

Dependent var.	Rural sample					Urban sample				
Vocabulary test z-score	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Siblings	0.031***					0.020				
	[0.006]					[0.013]				
Siblings*Female	-0.059***					-0.030**				
	[0.007]					[0.012]				
Brothers		0.016**					0.017			
		[0.006]					[0.016]			
Brothers*Female		-0.055***					-0.022			
		[0.009]					[0.014]			
Sisters		0.047***					0.023			
		[0.008]					[0.017]			
Sisters*Female		-0.070***					-0.036*			
		[0.010]					[0.020]			
Older siblings			0.018***					0.006		
			[0.007]					[0.018]		
Older siblings*Female			-0.014					-0.003		
			[0.009]					[0.018]		
Younger siblings			0.040***					0.030**		
			[0.007]					[0.013]		
Younger siblings*Female			-0.092***					-0.046***		
			[0.009]					[0.014]		
Older brothers				0.003	0.003				-0.021	0.131***
				[0.009]	[0.030]				[0.022]	[0.042]
Older brother*Female				-0.018	-0.031				0.029	-0.054
				[0.012]	[0.060]				[0.022]	[0.078]
Younger brothers				0.021**	0.022**				0.063***	0.063***
				[0.009]	[0.009]				[0.020]	[0.020]
Younger brothers*Female				-0.075***	-0.076***				-0.077***	-0.077***
				[0.013]	[0.013]				[0.021]	[0.021]
Older sisters				0.032***	0.033				0.032	0.186***
				[0.009]	[0.028]				[0.021]	[0.046]
Older sisters*Female				-0.011	-0.023				-0.035	-0.122
				[0.011]	[0.059]				[0.023]	[0.075]
Younger sisters				0.058***	0.058***				-0.004	-0.003
				[0.011]	[0.011]				[0.020]	[0.020]
Younger sisters*Female				-0.112***	-0.113***				-0.014	-0.015
				[0.014]	[0.014]				[0.028]	[0.029]
Firstborn vars	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Birth order vars	No	No	No	No	Yes	No	No	No	No	Yes
Observations	19,220	19,220	19,220	19,220	19,220	3,719	3,719	3,719	3,719	3,719
R-squared	0.405	0.406	0.407	0.408	0.409	0.335	0.335	0.336	0.338	0.340

1) The outcome variables are test z-scores with zero mean and a standard deviation
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in Table 2;
4) Column (5) includes birth order and its interaction term with female dummy instead of firstborn dummies.

Table 2-G4: OLS results - Math test

Dependent var.	**Panel A: Rural sample**				
Standardised Math test score	(1)	(2)	(3)	(4)	(5)
Siblings	-0.116***	-0.091***	-0.077***	0.040***	0.011***
	[0.006]	[0.006]	[0.005]	[0.006]	[0.003]
Siblings*Female		-0.049***	-0.053***	-0.083***	-0.009**
		[0.007]	[0.006]	[0.008]	[0.004]
Female	-0.356***	-0.213***	-0.212***	-0.246***	-0.062***
	[0.021]	[0.030]	[0.029]	[0.027]	[0.014]
Parental schooling years				0.064***	0.007***
				[0.003]	[0.001]
Han Chinese				0.089	0.032
				[0.064]	[0.023]
Birth order				-0.022***	-0.009**
				[0.007]	[0.004]
Birth order*Female				0.040***	0.005
				[0.009]	[0.005]
Age				-0.015***	-0.005**
				[0.005]	[0.002]
Age squared				0.000	0.000***
				[0.000]	[0.000]
Years of schooling					0.170***
					[0.001]
Birth county dummies	No	No	Yes	Yes	Yes
Birth county dummies*time	No	No	No	Yes	Yes
Observations	21,641	21,641	21,641	21,641	21,641
R-squared	0.099	0.102	0.213	0.37	0.801
Joint significance of Sibling and Sibling*Female (P value)		0.000	0.000	0.000	0.018
Dependent var.	**Panel B: Urban sample**				
Stadardized Math test score	(1)	(2)	(3)	(4)	(5)
Siblings	-0.112***	-0.105***	-0.093***	0.006	0.003
	[0.008]	[0.010]	[0.009]	[0.015]	[0.008]
Siblings*Female		-0.014	-0.012	-0.026	0.010
		[0.011]	[0.011]	[0.017]	[0.011]
female	-0.029	-0.001	-0.016	0.002	-0.019
	[0.019]	[0.033]	[0.030]	[0.029]	[0.031]
Parental schooling years				0.041***	-0.001
				[0.003]	[0.002]
Han Chinese				0.148*	0.069
				[0.083]	[0.047]
Birth order				0.008	0.009
				[0.017]	[0.007]
Birth order*Female				0	-0.012
				[0.018]	[0.009]
Age				0.316***	0.115***
				[0.023]	[0.014]
Age squared				-0.000***	0.000
				[0.000]	[0.000]
Years of schooling					0.161***
					[0.003]
Birth county dummies	No	No	Yes	Yes	Yes
Birth county dummies*time	No	No	No	Yes	Yes
Observations	4,058	4,058	4,058	4,058	4,058
R-squared	0.095	0.095	0.209	0.352	0.747
Joint significance of Sibling and Sibling*Female (P value)		0.000	0.000	0.196	0.131

1) Clustered by birth county for rural and prefecture for urban analysis. 2) Test year and county information source indicators are included in column (3)-(5).

Table 2-G5: IV results - Math test

	Rural sample		Urban sample	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Siblings	0.040***	0.276***	0.006	0.860**
	(0.006)	(0.040)	(0.015)	(0.388)
Siblings*Female	-0.083***	-0.236***	-0.026	-0.007
	(0.008)	(0.018)	(0.017)	(0.045)
Female	-0.246***	-0.133***	0.002	-0.070
	(0.027)	(0.036)	(0.029)	(0.067)
Observations	21641	21641	4058	4058
Adjusted R^2	0.361	0.304	0.313	-1.154

Note:
1) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in the main IV estimation in Table 4;
4) The instrument is the Birth Control Rate;

Table 2-G6: Proportion of women who prefer a son to a daughter

	Jilin	Shanghai	Shandong	Guangdong	Shaanxi	Ningxia	Total
Living in rural survey regions							
	48.3%	17.7%	55.7%	75.8%	68.4%	75.0%	55.7%
Sample size	1,494	1,560	1,480	1,475	1,531	1,486	9,026
Living in urban survey regions							
	28.7%	18.4%	22.4%	33.8%	21.8%	28.5%	21.6%
Sample size	1,475	1,560	1,470	1,457	1,529	1,491	8,982

[1)] Source: the Sampling Survey Data of Women's Status in Contemporary China in 1992

Table 2-G7: IV results by province

	Rural sample			
	Shanghai	Jilin	Shandong	Shaanxi
	(1)	(2)	(3)	(4)
Siblings	0.040***	0.276***	0.006	0.860**
	(0.006)	(0.040)	(0.015)	(0.388)
Siblings*Female	-0.083***	-0.236***	-0.026	-0.007
	(0.008)	(0.018)	(0.017)	(0.045)
Female	-0.246***	-0.133***	0.002	-0.070
	(0.027)	(0.036)	(0.029)	(0.067)
Observations	21641	21641	4058	4058
Adjusted R^2	0.361	0.304	0.313	-1.154
	Urban sample			
	Shanghai	Jilin	Shandong	Shaanxi
	(1)	(2)	(3)	(4)
Siblings	0.040***	0.276***	0.006	0.860**
	(0.006)	(0.040)	(0.015)	(0.388)
Siblings*Female	-0.083***	-0.236***	-0.026	-0.007
	(0.008)	(0.018)	(0.017)	(0.045)
Female	-0.246***	-0.133***	0.002	-0.070
	(0.027)	(0.036)	(0.029)	(0.067)
Observations	21641	21641	4058	4058
Adjusted R^2	0.361	0.304	0.313	-1.154

Note:
1) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in the main IV estimation in Table 4;
4) The instrument is the Birth Control Rate;

Table 2-G8: IV results using different year average Birth Control Rate

	Rural sample			
	At one's birth	within 3 years	within 5 years	within 7 years
	(1)	(2)	(3)	(4)
Siblings	0.130***	0.123***	0.125***	0.126***
	(0.041)	(0.040)	(0.038)	(0.037)
Siblings*Female	-0.330***	-0.332***	-0.331***	-0.332***
	(0.017)	(0.017)	(0.017)	(0.017)
Female	0.056	0.060*	0.059*	0.060*
	(0.036)	(0.036)	(0.036)	(0.035)
Observations	21641	21641	21641	21641
Adjusted R^2	0.375	0.374	0.374	0.374
Kleibergen-Paap Wald rk F statistics	136.134	142.592	148.736	155.207
	Urban sample			
	At one's birth	within 3 years	within 5 years	within 7 years
	(1)	(2)	(3)	(4)
Siblings	0.170	0.168	0.174	0.160
	(0.194)	(0.174)	(0.164)	(0.152)
Siblings*Female	-0.102***	-0.101***	-0.099***	-0.098***
	(0.028)	(0.027)	(0.028)	(0.027)
Female	-0.018	-0.018	-0.018	-0.017
	(0.033)	(0.032)	(0.033)	(0.032)
Observations	4058	4058	4058	4058
Adjusted R^2	0.262	0.264	0.260	0.267
Kleibergen-Paap Wald rk F statistics	3.546	3.692	3.383	3.426

Note:
1) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in the main IV estimation in Table 4;
4) 'At one's birth' means the Birth Control Rate at one's birth year is used as the IV. 'within 3 years' means the 3-year average Birth Control Rate, 'within 5 years' the 5-year average, and 'within 7 years' the 7-year average is used as the IV

Table 2-G9: Birth order group dummy variables added

	Rural sample		Urban sample	
	OLS	IV	OLS	IV
Siblings	0.041***	0.121***	0.034***	0.167
	[0.007]	[0.038]	[0.013]	[0.151]
Sibling*Female	-0.096***	-0.323***	-0.056***	-0.101***
	[0.009]	[0.017]	[0.014]	[0.028]
Female	-0.102***	0.323***	0.038	0.066*
	[0.028]	[0.041]	[0.026]	[0.040]
Birth order=2	-0.066***	-0.121***	-0.032	-0.143
	[0.017]	[0.027]	[0.030]	[0.117]
Birth order=3	-0.017	-0.151***	-0.097	-0.298
	[0.022]	[0.047]	[0.063]	[0.217]
Birth order=4	-0.055**	-0.245***	-0.096*	-0.399
	[0.027]	[0.071]	[0.052]	[0.323]
Birth order=5	-0.077**	-0.345***	-0.078	-0.491
	[0.035]	[0.100]	[0.074]	[0.451]
Birth order=6	-0.137***	-0.468***	-0.150	-0.671
	[0.048]	[0.134]	[0.101]	[0.597]
Birth order=7	-0.143*	-0.539***	-0.198	-0.829
	[0.072]	[0.172]	[0.143]	[0.688]
Birth order=8	-0.304***	-0.771***	-0.378***	-1.185
	[0.103]	[0.204]	[0.106]	[0.923]
(Birth order=2)*Female	0.001	0.143***	0.040	0.099**
	[0.025]	[0.029]	[0.047]	[0.051]
(Birth order=3)*Female	-0.016	0.311***	0.140**	0.235**
	[0.032]	[0.038]	[0.065]	[0.092]
(Birth order=4)*Female	0.060	0.554***	0.125	0.256***
	[0.041]	[0.050]	[0.076]	[0.093]
(Birth order=5)*Female	0.132**	0.814***	0.174	0.338**
	[0.051]	[0.060]	[0.108]	[0.151]
(Birth order=6)*Female	0.259***	1.122***	0.253**	0.447***
	[0.066]	[0.083]	[0.115]	[0.162]
(Birth order=7)*Female	0.349***	1.396***	0.355**	0.614***
	[0.096]	[0.119]	[0.156]	[0.158]
(Birth order=8)*Female	0.705***	1.969***	0.249	0.517**
	[0.147]	[0.166]	[0.204]	[0.262]
Observations	21641	21641	4058	4058
F-test		156.20		3.626
Joint significance of Sibling and Sibling*Female (P value)	0.000	0.000	0.001	0.001

Note:
1) The estimates for 9th-12th birth order group are excluded because each group has less than 50 observations, which may be misleading;
2) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
3) Clustered by birth county for rural and prefecture for urban sample;
4) Controls include the same set of covariates in Table 4 excluding birth order and its interaction with female dummy;
5) The 7-year average Birth control rate is used as the instrument;
6) Birth order=# is a dummy variable indicating one's birth order is equal to #

Table 2-G10: IV results including years of schooling

	Rural sample		Urban sample	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Siblings	0.020***	-0.047*	0.031***	0.078
	(0.005)	(0.028)	(0.009)	(0.121)
Siblings*Female	-0.037***	-0.168***	-0.032***	-0.066***
	(0.006)	(0.013)	(0.012)	(0.022)
Female	-0.045***	0.117***	-0.023	-0.024
	(0.016)	(0.024)	(0.028)	(0.031)
Years of schooling	0.133***	0.130***	0.103***	0.103***
	(0.002)	(0.002)	(0.003)	(0.004)
Observations	21641	21641	4058	4058
Adjusted R^2	0.683	0.640	0.516	0.513
Kleibergen-Paap Wald rk F statistics		157.08		3.49
P value for Joint significance test				
Sibling and Sibling*Female	0.000	0.000	0.002	0.010

Note:
1) The outcome variables are vocabulary test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in the main IV estimation in Table 4;

Table 2-G11: Son preference mechanism analysis

Dependent variable: Coefficient on the interaction term between the number of siblings and female dummy						
	Panel A: 19 provinces included					
	Q1	Q2	Q3	Q4	Q5	Q6
The proportion of people agreed	-0.387**	-0.244	0.026	0.033	-0.076	0.105
	[0.143]	[0.173]	[0.074]	[0.071]	[0.077]	[0.087]
Observations	19	19	19	19	19	19
R-squared	0.301	0.105	0.007	0.013	0.054	0.079
	Panel B: All 25 provinces included					
	Q1	Q2	Q3	Q4	Q5	Q6
The proportion of people agreed	-0.282*	-0.115	0.066	0.075	-0.077	0.142
	[0.163]	[0.177]	[0.094]	[0.115]	[0.071]	[0.095]
Observations	25	25	25	25	25	25
R-squared	0.115	0.018	0.021	0.018	0.049	0.088

1) The outcome variable is the estimated coefficient on the interaction term between the number of siblings and female dummy at provincial level. We run the main IV estimation (column 2 in Table 5) for each province and estimate the coefficients;
2) The independent variables are the proportion of people who agreed to each question at provincial level - Q1 "Women should give birth to at least one boy to continue family lineage", Q2 "A man should live with his parents after marriage", Q3 "Men should focus on career, while women should focus on family", Q4 "Marrying well is more important for women than doing well", Q5 "Women should have ate least one child", Q6 "Men should do half of the housework";
3) 19 provinces: Individual observations in the IV estimation range from 400 to 2755. 25 provinces: Individual observations in the IV estimation ranges from 44 to 2755 (See Section 6.1 for more details);

Table 2-G12: Vocabulary test score robustness check

	Panel A: IV results									
	Rural sample					Urban sample				
	Voca	Max score	Raw score	2010 only	2014 only	Voca	Max score	Raw score	2010 only	2014 only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Siblings	0.126***	0.166***	1.381***	0.153***	0.099**	0.160	0.131	1.764	0.167	0.070
	(0.037)	(0.039)	(0.410)	(0.040)	(0.043)	(0.152)	(0.145)	(1.668)	(0.170)	(0.171)
Siblings*Female	-0.332***	-0.357***	-3.645***	-0.316***	-0.334***	-0.098***	-0.093***	-1.079***	-0.111***	-0.118***
	(0.017)	(0.018)	(0.187)	(0.019)	(0.019)	(0.027)	(0.027)	(0.301)	(0.031)	(0.038)
Female	0.060*	0.055	0.658*	0.014	0.094**	-0.017	-0.027	-0.189	-0.020	-0.008
	(0.035)	(0.036)	(0.388)	(0.038)	(0.038)	(0.032)	(0.033)	(0.354)	(0.041)	(0.036)
Observations	21641	21641	21641	20401	16468	4058	4058	4058	3869	2501
Adjusted R^2	0.374	0.352	0.374	0.347	0.343	0.267	0.256	0.266	0.233	0.290
	Panel B: OLS results									
	Rural sample					Urban sample				
	Voca	Max score	Raw score	2010 only	2014 only	Voca	Max score	Raw score	2010 only	2014 only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Siblings	0.042***	0.044***	0.462***	0.045***	0.035***	0.034***	0.029**	0.367***	0.042***	0.028
	(0.007)	(0.007)	(0.072)	(0.007)	(0.008)	(0.012)	(0.012)	(0.136)	(0.016)	(0.018)
Siblings*Female	-0.095***	-0.105***	-1.040***	-0.096***	-0.084***	-0.055***	-0.056***	-0.604***	-0.058***	-0.055***
	(0.009)	(0.009)	(0.096)	(0.009)	(0.010)	(0.014)	(0.014)	(0.151)	(0.015)	(0.021)
Female	-0.188***	-0.200***	-2.068***	-0.221***	-0.195***	-0.010	-0.021	-0.106	-0.016	-0.016
	(0.026)	(0.027)	(0.290)	(0.028)	(0.029)	(0.029)	(0.031)	(0.319)	(0.036)	(0.034)
Observations	21641	21641	21641	20401	16468	4058	4058	4058	3869	2501
Adjusted R^2	0.410	0.389	0.411	0.374	0.384	0.296	0.275	0.295	0.258	0.294

Note: 1) 'Main' means the outcome variable is individual average standardised vocabulary score between 2010 and 2014; 'Max' means the outcome variable is individual maximum standardised vocabulary score between 2010 and 2014; 'Raw' means the outcome variable is individual average raw vocabulary score between 2010 and 2014; '2010 only' means the outcome variable is 2010 test standardised score; '2014 only' means the outcome variable is 2014 test standardised score;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Controls include the same set of covariates in the main results in Table 4;
4) Instrument : Birth Control Rate 7 year average;

Table 2-G13: Gender gap in NCEE performance

	NCEE			**CFPS**
	OLS	OLS	IV	IV (Main results)
Siblings	0.072***	0.048**	-0.323	0.160
	[0.017]	[0.021]	[0.366]	[0.152]
Sibling*Female	-0.069***	-0.056*	-0.125***	-0.098***
	[0.023]	[0.031]	[0.048]	[0.027]
Female	-0.031	-0.070	0.151	-0.017
	[0.091]	[0.114]	[0.157]	[0.032]
Observations	4,753	3,223	3,223	4,058
Kleibergen-Paap Wald rk F statistics			1.381	3.43
Joint test	0.076	0.029	0.020	0.001

Note:
1) The outcome variable is the standardised NCEE test score in column 1-3;
2) Column 4 presents the main IV results in Table 4 for urban sample;
3) Control variables include the same set of covariates in the main analysis in Table 4 in addition to test year and test province dummy variables;
4) The instrument in column 3 is the 7 year average Birth Control Rate;

Table 2-G14: IV results

Dependent var.	Rural			Urban		
Vocabulary test z-score	(OLS)	(IV) Birth control rate	(IV) Policy fertility	(OLS)	(IV) Birth control rate	(IV) Policy fertility
Siblings	0.042***	0.126***	-0.003	0.034***	0.160	-3.199
	[0.007]	[0.037]	[0.081]	[0.012]	[0.152]	[157.191]
Siblings*Female	-0.095***	-0.332***	-0.342***	-0.055***	-0.098***	-0.093
	[0.009]	[0.017]	[0.025]	[0.014]	[0.027]	[1.502]
Female	-0.188***	0.060*	0.105**	-0.010	-0.017	0.259
	[0.026]	[0.035]	[0.046]	[0.029]	[0.032]	[13.066]
Parental schooling years	0.050***	0.051***	0.052***	0.035***	0.036***	-0.004
	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[1.889]
Han Chinese	0.120*	0.119*	0.114*	0.061	0.063	0.061
	[0.064]	[0.064]	[0.062]	[0.065]	[0.067]	[0.293]
Birth order	-0.021***	-0.091***	-0.005	-0.031**	-0.131	2.413
	[0.006]	[0.025]	[0.053]	[0.014]	[0.116]	[118.953]
Birth order*Female	0.046***	0.227***	0.230***	0.049***	0.088***	0.137
	[0.009]	[0.013]	[0.019]	[0.016]	[0.027]	[4.000]
Age	-0.002	0.002	0.018*	0.447***	0.495***	-1.067
	[0.005]	[0.007]	[0.011]	[0.017]	[0.073]	[73.586]
Age squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000*	0.004
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.187]
Observations	21,641	21,641	21,641	4,058	4,058	4,058
Kleibergen-Paap Wald rk F statistics		155.21	18.74		3.43	0.00
P value for Joint significance test						
Sibling and Sibling*Female	0.000	0.000	0.000	0.001	0.001	0.941

Note:
1) The outcome variables are test z-scores with zero mean and a standard deviation;
2) Clustered by birth county for rural and prefecture for urban sample;
3) Control variables include the same set of covariates in Table 3;
4) The instrument is 7-year average Birth control rate;

Table 2-G15: First stage results

	Rural sample			
	(1)	(2)	(3)	(4)
Dependent var.	Siblings	Siblings*Female	Siblings	Siblings*Female
Birth Control Rate (%)	-0.013***	0.002***		
	[0.001]	[0.001]		
Birth Control Rate (%)*Female	0.002***	-0.017***		
	[0.000]	[0.001]		
Policy Fertility Index			0.139***	-0.164***
			[0.022]	[0.013]
Policy Fertility Index*Female			-0.042**	0.439***
			[0.017]	[0.019]
Controls	Yes	Yes	Yes	Yes
Observations	21,641	21,641	21,641	21,641
Kleibergen-Paap Wald rk F statistics		155.21		18.74
	Urban sample			
	(1)	(2)	(3)	(4)
Dependent var.	Siblings	Siblings*Female	Siblings	Siblings*Female
Birth Control Rate	-0.006***	0.006***		
	[0.002]	[0.001]		
Birth Control Rate*Female	0.001	-0.019***		
	[0.001]	[0.001]		
Policy Fertility Index			0.003	-0.204***
			[0.049]	[0.020]
Policy Fertility Index*Female			-0.004	0.455***
			[0.026]	[0.028]
Controls	Yes	Yes	Yes	Yes
Observations	4,058	4,058	4,058	4,058
Kleibergen-Paap Wald rk F statistics		3.43		0.00

Note:
1) Robust std. err. are clustered by birth county for rural and prefecture for urban analysis.
2) Controls include birth county (or prefecture) fixed effects, age, age square term, birth order, an interaction of birth order with female variable, Han Chinese indicator, parental mean schooling years, county information source indicators, test year indicators, and an interaction between time linear trend variable*birth region fixed effect.

APPENDIX H : NCEE details

Table 2-H1: Main sample selection process

Description	N
Total number of individuals in UREES	66,268
College entrance exam score reported	6,005
General college entrance exam takers	5,466
- Test year information available	5,434
- Attended the NCEE since 1977	5,228
- Urban household registration	5,149
- Sibling information available	5,099
- Attended in the surveyed provinces	4,857

Table 2-H2: Test procedure details by province

Province (code)	Own Test papers used since	standardised scores provided
Beijing (11)	2002	n/a
Shanxi (14)	n/a	n/a
Liaoning (21)	2004	n/a
Heilongjiang (23)	n/a	n/a
Zhejiang (33)	2004	n/a
Anhui (34)	2004	n/a
Hubei (42)	2004	n/a
Guangdong (44)	2004	1988-2006
Sichuan (51)	2004	n/a
Guizhou (52)	n/a	n/a
Shaanxi (61)	2004	1994-2001
Gansu (62)	n/a	n/a

[1)] Since our sample include those who attended the NCEE between 1977 and 2004 only, we did not collect timing information when the province started to use their own paper since 2005.

Appendix to Chapter 3

Appendix A: Figures and Tables

Table 3-A1: Main sample generation

Description	Observations
All CFPS respondents in 2010	33600
+ Born between 1950 and 1990	25243
++ Height info. available	25095
+++ *hukou* at age 3 info. available	24793
++++ Birth county info. available	19430
+++++ Birth county gazetteer data available	18715
++++++ Individual gender and ethnicity info. available	18677
+++++++ Sibling info. available	18511
++++++++ Height>190cm excluded	18508
Final sample - Rural	16051
Final sample - Urban	2457

Table 3-A2: Main estimation excluding 1950s cohort — Rural sample

	Main sample			Excluding 1950s birth cohort		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Agricultural output p.c.- age 0 to 3	0.427**	0.746***	0.074	0.201	0.262	0.051
	(0.168)	(0.233)	(0.217)	(0.181)	(0.268)	(0.285)
Log Industrial output p.c.- age 0 to 3	0.206***	0.259**	0.158	0.155*	0.264**	0.100
	(0.077)	(0.101)	(0.103)	(0.091)	(0.105)	(0.148)
Log Gov. social expenditure p.c.- age 0 to 3	0.477***	0.646***	0.332***	0.669***	1.089***	0.300*
	(0.099)	(0.136)	(0.115)	(0.123)	(0.192)	(0.162)
Death rate (‰) - age 0 to 3	-0.032**	-0.015	-0.053***	-0.025	0.002	-0.050**
	(0.012)	(0.015)	(0.016)	(0.017)	(0.024)	(0.020)
Parental schooling years	0.077***	0.102***	0.049**	0.049***	0.066***	0.031
	(0.016)	(0.023)	(0.022)	(0.017)	(0.025)	(0.024)
Father's occupation rank at child's age 14	0.002***	0.002***	0.002**	0.002***	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Father's party membership at child's age 14	0.360***	0.299	0.382**	0.310**	0.292	0.256
	(0.129)	(0.195)	(0.190)	(0.143)	(0.220)	(0.213)
Male	10.022***			10.123***		
	(0.096)			(0.110)		
Han Chinese	0.108	0.259	-0.005	0.232	0.211	0.369
	(0.250)	(0.365)	(0.329)	(0.292)	(0.397)	(0.338)
Number of siblings	-0.085**	-0.071	-0.055	-0.132***	-0.138**	-0.065
	(0.039)	(0.047)	(0.053)	(0.047)	(0.063)	(0.061)
Birth order	0.061	0.034	0.053	0.046	0.053	-0.008
	(0.037)	(0.050)	(0.050)	(0.044)	(0.062)	(0.059)
Observations	16051	8098	7953	12331	6161	6170
Adjusted R^2	0.560	0.225	0.164	0.569	0.235	0.159

Note:

1) Rural sample only

2) Robust Standard Errors (SEs) are presented in parentheses;

3) SEs are clustered at the county levels;

4) Unit: log of 10000 yuan; measured in 1990 price level

5) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;

6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A3: Estimated results with total output

	Rural sample			Urban sample		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Total output p.c. - age 0 to 3	0.520***	0.812***	0.228	0.185	0.046	0.309*
	(0.145)	(0.197)	(0.188)	(0.131)	(0.156)	(0.183)
Log Gov. social expenditure p.c.- age 0 to 3	0.546***	0.724***	0.386***	0.874***	1.099***	0.532***
	(0.097)	(0.128)	(0.117)	(0.087)	(0.117)	(0.162)
Death rate (‰) - age 0 to 3	-0.035***	-0.018	-0.054***	0.021	0.019	0.019
	(0.013)	(0.016)	(0.016)	(0.037)	(0.046)	(0.051)
Parental schooling years	0.079***	0.106***	0.050**	0.125***	0.172***	0.114***
	(0.016)	(0.024)	(0.022)	(0.033)	(0.038)	(0.043)
Father's occupation rank at child's age 14	0.002***	0.002**	0.002**	0.000	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Father's party membership at child's age 14	0.354***	0.284	0.382**	-0.091	-0.361	0.202
	(0.130)	(0.195)	(0.190)	(0.288)	(0.408)	(0.346)
Male	10.023***			11.339***		
	(0.096)			(0.242)		
Han Chinese	0.115	0.287	-0.009	-0.449	0.083	-1.121
	(0.249)	(0.360)	(0.329)	(0.847)	(0.844)	(1.017)
Number of siblings	-0.088**	-0.076	-0.056	-0.192*	-0.135	-0.257*
	(0.039)	(0.047)	(0.053)	(0.107)	(0.179)	(0.146)
Birth order	0.064*	0.038	0.056	-0.061	-0.049	-0.051
	(0.037)	(0.050)	(0.050)	(0.122)	(0.222)	(0.177)
Observations	16051	8098	7953	2457	1270	1187
Adjusted R^2	0.560	0.224	0.164	0.610	0.218	0.120

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples, respectively;

3) Unit: log of 10000 yuan; measured in 1990 price level

4) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;

5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A4: Selected results with Interaction terms

	Rural sample	Urban sample
	(1)	(2)
(a) Log Agricultural output p.c.- age 0 to 3	-0.021	-0.198***
	(0.194)	(0.068)
(b) Log Industrial output p.c.- age 0 to 3	0.100	-0.030
	(0.089)	(0.116)
(c) Log Gov. social expenditure p.c.- age 0 to 3	0.349***	0.174
	(0.108)	(0.165)
(d) Death rate (‰) - age 0 to 3	-0.024*	0.003
	(0.013)	(0.030)
Exposure	2.620**	6.713***
	(1.059)	(1.542)
(e) Exposure × Log Agricultural output p.c.- age 0 to 3	0.211	-0.097
	(0.239)	(0.111)
(f) Exposure × Log Industrial output p.c.- age 0 to 3	0.124	0.254
	(0.090)	(0.157)
(g) Exposure × Log Gov. social expenditure p.c.- age 0 to 3	-0.048	0.595***
	(0.164)	(0.202)
(h) Exposure × Death rate (‰) - age 0 to 3	-0.096**	-0.093
	(0.043)	(0.110)
Observations	16051	2457
Adjusted R^2	0.561	0.615
Total effect for the exposed group		
Coefficient (a)+(e)	0.190	-0.295
Test h_0 : Coefficient on $(a)+(e)=0$ [P value]	[0.390]	[0.004]
Coefficient (b)+(f)	0.224	0.224
Test h_0 : Coefficient on $(b)+(f)=0$ [P value]	[0.057]	[0.062]
Coefficient on (c)+(g)	0.301	0.769
Test h_0 : Coefficient on $(c)+(g)=0$ [P value]	[0.051]	[0.000]
Coefficient on (d)+(h)	-0.120	-0.090
Test h_0 : Coefficient on $(d)+(h)=0$ [P value]	[0.004]	[0.430]

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples;

3) Unit: log of 10000 yuan; measured in 1990 price level;

4) All the covariates included in Table 3 are included;

5) * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 3-A5: IV first stage results

	Rural			Urban		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Birth Control Rate (%)	-0.015***	-0.015***	-0.015***	-0.013***	-0.013***	-0.014***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Observations	16049	8097	7952	2457	1270	1187

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples;

3) The dependent variable is the number of siblings;

4) All the variables in Table 3-7 are included;

5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A6: Main estimation excluding predicted observations (Excluding one missing variable at a time)

	Main sample			Excluding predicted observations of: Agricultural output			Industrial output			Death rate			Government social expenditure		
	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Log Agr. output p.c.- age 0 to 3	0.427**	0.746***	0.074	0.393*	0.768**	0.042	0.422*	0.722**	0.118	0.329*	0.763***	-0.152	0.367*	0.737***	-0.012
	(0.168)	(0.233)	(0.217)	(0.224)	(0.316)	(0.307)	(0.242)	(0.315)	(0.310)	(0.188)	(0.272)	(0.225)	(0.192)	(0.275)	(0.250)
Log Ind. output p.c.- age 0 to 3	0.206***	0.259**	0.158	0.209*	0.331**	0.105	0.220**	0.250*	0.204	0.180**	0.198*	0.170	0.062	0.186	-0.045
	(0.077)	(0.101)	(0.103)	(0.110)	(0.154)	(0.147)	(0.099)	(0.134)	(0.132)	(0.086)	(0.115)	(0.117)	(0.097)	(0.138)	(0.121)
Log Gov. social exp. p.c.- age 0 to 3	0.477***	0.646***	0.332***	0.481***	0.706***	0.249	0.449***	0.724***	0.186	0.593***	0.823***	0.409***	0.673***	0.915***	0.472***
	(0.099)	(0.136)	(0.115)	(0.152)	(0.203)	(0.174)	(0.156)	(0.213)	(0.170)	(0.129)	(0.169)	(0.142)	(0.117)	(0.150)	(0.147)
Death rate (‰) - age 0 to 3	-0.032**	-0.015	-0.053***	-0.008	0.006	-0.028	-0.012	0.004	-0.033*	-0.042***	-0.026	-0.061***	-0.015	0.013	-0.047***
	(0.012)	(0.015)	(0.016)	(0.015)	(0.019)	(0.021)	(0.017)	(0.022)	(0.019)	(0.014)	(0.017)	(0.018)	(0.013)	(0.016)	(0.016)
Parental schooling years	0.077***	0.102***	0.049**	0.081***	0.087***	0.075**	0.085***	0.120***	0.044	0.064***	0.074***	0.051**	0.085***	0.095***	0.069***
	(0.016)	(0.023)	(0.022)	(0.022)	(0.030)	(0.030)	(0.021)	(0.031)	(0.028)	(0.018)	(0.025)	(0.024)	(0.018)	(0.027)	(0.025)
Father's occupation rank at child's age 14	0.002***	0.002***	0.002**	0.003***	0.003***	0.002*	0.001*	0.002*	0.001	0.002***	0.002***	0.002**	0.002***	0.002**	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Father's party membership at child's age 14	0.360***	0.299	0.382**	0.357*	0.280	0.384	0.423**	0.328	0.463*	0.371***	0.371*	0.326*	0.451***	0.411*	0.433*
	(0.129)	(0.195)	(0.190)	(0.189)	(0.303)	(0.284)	(0.165)	(0.280)	(0.246)	(0.135)	(0.221)	(0.196)	(0.154)	(0.241)	(0.228)
Male	10.022***			10.164***			10.044***			9.987***			9.994***		
	(0.096)			(0.126)			(0.126)			(0.098)			(0.110)		
Han Chinese	0.108	0.259	-0.005	0.091	-0.200	0.483	0.070	-0.094	0.298	0.005	0.043	0.019	0.309	0.339	0.366
	(0.250)	(0.365)	(0.329)	(0.289)	(0.500)	(0.457)	(0.354)	(0.488)	(0.420)	(0.280)	(0.396)	(0.365)	(0.275)	(0.359)	(0.383)
Number of siblings	-0.085**	-0.071	-0.055	-0.061	-0.033	-0.059	-0.091*	-0.072	-0.076	-0.087**	-0.059	-0.079	-0.076*	-0.050	-0.059
	(0.039)	(0.047)	(0.053)	(0.053)	(0.066)	(0.074)	(0.051)	(0.061)	(0.071)	(0.044)	(0.055)	(0.056)	(0.045)	(0.054)	(0.063)
Birth order	0.061	0.034	0.053	0.025	0.015	0.016	0.072	0.069	0.047	0.056	0.017	0.069	0.060	0.049	0.036
	(0.037)	(0.050)	(0.050)	(0.051)	(0.078)	(0.066)	(0.045)	(0.068)	(0.061)	(0.040)	(0.055)	(0.053)	(0.043)	(0.057)	(0.058)
Observations	16051	8098	7953	8480	4227	4253	9209	4650	4559	13462	6781	6681	12043	6087	5956
Adjusted R^2	0.560	0.225	0.164	0.567	0.213	0.158	0.562	0.238	0.168	0.559	0.218	0.164	0.560	0.237	0.169

Note:

1) Rural sample only

2) Robust Standard Errors (SEs) are presented in parentheses;

3) SEs are clustered at the county and prefecture levels for the rural and urban samples;

4) Unit: log of 10000 yuan; measured in 1990 price level

5) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation;

6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A7: Main estimation excluding predicted observations (Excluding all missing variables at once)

	Main sample			All predicted values are excluded								
	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log Agricultural output p.c.- age 0 to 3	0.427**	0.746***	0.074	0.046	0.344	-0.310	-0.222	-0.090	-0.435	-0.230	-0.089	-0.446
	(0.168)	(0.233)	(0.217)	(0.273)	(0.413)	(0.342)	(0.275)	(0.433)	(0.377)	(0.275)	(0.431)	(0.378)
Log Industrial output p.c.- age 0 to 3	0.206***	0.259**	0.158	0.148	0.325*	0.012	0.006	0.097	-0.054	0.021	0.099	-0.037
	(0.077)	(0.101)	(0.103)	(0.113)	(0.173)	(0.149)	(0.119)	(0.184)	(0.152)	(0.118)	(0.178)	(0.154)
Log Gov. social expenditure p.c.- age 0 to 3	0.477***	0.646***	0.332***	0.881***	1.088***	0.674***	0.608***	0.649**	0.546**	0.546***	0.576**	0.507*
	(0.099)	(0.136)	(0.115)	(0.208)	(0.270)	(0.239)	(0.204)	(0.258)	(0.259)	(0.203)	(0.257)	(0.262)
Death rate (‰) - age 0 to 3	-0.032**	-0.015	-0.053***	-0.010	0.014	-0.039	0.010	0.047*	-0.031	-0.001	0.036	-0.039
	(0.012)	(0.015)	(0.016)	(0.022)	(0.026)	(0.032)	(0.023)	(0.027)	(0.032)	(0.022)	(0.025)	(0.033)
Parental schooling years	0.077***	0.102***	0.049**	0.067***	0.066*	0.074**	0.055**	0.053	0.067*	0.052**	0.050	0.063*
	(0.016)	(0.023)	(0.022)	(0.025)	(0.037)	(0.035)	(0.026)	(0.037)	(0.035)	(0.025)	(0.036)	(0.035)
Father's occupation rank at child's age 14	0.002***	0.002***	0.002**	0.003***	0.003**	0.002	0.003***	0.003**	0.002	0.002***	0.003**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Father's party membership at child's age 14	0.360***	0.299	0.382**	0.587***	0.451	0.629**	0.615***	0.478	0.649**	0.636***	0.516	0.655**
	(0.129)	(0.195)	(0.190)	(0.205)	(0.378)	(0.308)	(0.205)	(0.375)	(0.304)	(0.207)	(0.375)	(0.304)
Male	10.022***			10.289***			10.261***			10.242***		
	(0.096)			(0.145)			(0.145)			(0.144)		
Han Chinese	0.108	0.259	-0.005	-0.097	-0.383	0.195	-0.078	-0.412	0.222	-0.086	-0.405	0.206
	(0.250)	(0.365)	(0.329)	(0.317)	(0.507)	(0.459)	(0.320)	(0.520)	(0.461)	(0.325)	(0.512)	(0.471)
Number of siblings	-0.085**	-0.071	-0.055	-0.094	-0.079	-0.079	-0.056	-0.006	-0.064	0.026	0.072	-0.007
	(0.039)	(0.047)	(0.053)	(0.075)	(0.089)	(0.104)	(0.072)	(0.090)	(0.102)	(0.077)	(0.089)	(0.119)
Birth order	0.061	0.034	0.053	0.079	0.059	0.090	0.066	0.030	0.085	0.090	0.068	0.095
	(0.037)	(0.050)	(0.050)	(0.067)	(0.094)	(0.084)	(0.067)	(0.095)	(0.083)	(0.065)	(0.096)	(0.082)
Exposure							1.447***	2.344***	0.677	2.327***	3.248***	1.287**
							(0.429)	(0.755)	(0.525)	(0.430)	(0.809)	(0.602)
Exposure × Number of siblings										-0.422***	-0.452**	-0.283*
										(0.120)	(0.185)	(0.152)
Observations	16051	8098	7953	5879	2923	2956	5879	2923	2956	5879	2923	2956
Adjusted R^2	0.560	0.225	0.164	0.583	0.231	0.176	0.584	0.235	0.176	0.585	0.236	0.177

Note:

1) Rural sample only

2) Robust Standard Errors (SEs) are presented in parentheses;

3) SEs are clustered at the county and prefecture levels for the rural and urban samples;

4) Unit: log of 10000 yuan; measured in 1990 price level

5) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;

6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A8: Estimated results with parental height variable included

	without parental height			with parental height		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Agricultural output p.c.- age 0 to 3	0.950**	0.969*	1.007	0.876**	0.975**	0.473
	(0.414)	(0.494)	(1.136)	(0.383)	(0.465)	(1.069)
Log Industrial output p.c.- age 0 to 3	0.042	0.008	0.433	-0.050	-0.056	0.182
	(0.212)	(0.241)	(0.542)	(0.210)	(0.234)	(0.544)
Log Gov. social expenditure p.c.- age 0 to 3	0.802***	1.070***	-0.278	0.578**	0.789**	-0.136
	(0.281)	(0.350)	(0.558)	(0.271)	(0.336)	(0.538)
Death rate (‰) - age 0 to 3	-0.010	0.020	0.033	-0.008	0.023	-0.011
	(0.041)	(0.042)	(0.096)	(0.039)	(0.042)	(0.082)
Parental schooling years	0.003	-0.002	0.032	-0.024	-0.024	-0.006
	(0.034)	(0.040)	(0.090)	(0.033)	(0.039)	(0.082)
Father's occupation rank at child's age 14	0.002*	0.002	0.003	0.001	0.001	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Father's party membership at child's age 14	0.619**	0.775**	0.022	0.548*	0.663**	0.060
	(0.296)	(0.322)	(0.707)	(0.290)	(0.326)	(0.672)
Male	11.025***			11.046***		
	(0.266)			(0.263)		
Han Chinese	0.286	0.707	-1.186	0.385	0.801	-0.962
	(0.501)	(0.607)	(1.000)	(0.462)	(0.599)	(0.976)
Number of siblings	-0.249***	-0.273***	0.123	-0.231**	-0.263**	0.144
	(0.095)	(0.105)	(0.243)	(0.101)	(0.109)	(0.206)
Birth order	0.079	0.185	-0.359	0.149	0.262**	-0.302
	(0.102)	(0.124)	(0.241)	(0.106)	(0.127)	(0.229)
Mother's height (cm)				0.296***	0.273***	0.372***
				(0.024)	(0.028)	(0.048)
Observations	3141	2360	781	3141	2360	781
Adjusted R^2	0.507	0.254	0.173	0.553	0.313	0.298

Note:

1) Rural sample only

2) Robust Standard Errors (SEs) are presented in parentheses;

3) SEs are clustered at the county and prefecture levels for the rural and urban samples;

4) Unit: log of 10000 yuan; measured in 1990 price level

5) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;

6) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A9: Main estimation with three decade dummies

	Rural sample			Urban sample		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Agricultural output p.c.- age 0 to 3	0.253	0.359	0.101	-0.171**	0.043	-0.375***
	(0.169)	(0.251)	(0.241)	(0.081)	(0.100)	(0.136)
Log Industrial output p.c.- age 0 to 3	0.157*	0.164	0.163	0.031	-0.115	0.156
	(0.080)	(0.104)	(0.105)	(0.112)	(0.192)	(0.095)
Log Gov. social expenditure p.c.- age 0 to 3	0.193*	0.332**	0.077	0.365***	0.702***	-0.034
	(0.098)	(0.152)	(0.126)	(0.114)	(0.143)	(0.239)
Death rate (‰) - age 0 to 3	-0.014	0.004	-0.035**	0.032	0.024	0.027
	(0.012)	(0.014)	(0.017)	(0.038)	(0.050)	(0.055)
Parental schooling years	0.062***	0.085***	0.039*	0.101***	0.153***	0.096**
	(0.016)	(0.023)	(0.023)	(0.032)	(0.042)	(0.041)
Father's occupation rank at child's age 14	0.002***	0.002***	0.002**	0.001	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Father's party membership at child's age 14	0.376***	0.325	0.381**	-0.000	-0.258	0.288
	(0.131)	(0.197)	(0.190)	(0.303)	(0.403)	(0.368)
Male	10.030***			11.310***		
	(0.097)			(0.235)		
Han Chinese	0.105	0.241	-0.004	-0.413	0.128	-1.030
	(0.247)	(0.361)	(0.328)	(0.837)	(0.823)	(1.047)
Number of siblings	-0.067*	-0.038	-0.052	-0.123	-0.054	-0.219
	(0.039)	(0.049)	(0.053)	(0.105)	(0.178)	(0.142)
Birth order	0.052	0.024	0.046	-0.062	-0.027	-0.057
	(0.038)	(0.052)	(0.050)	(0.121)	(0.222)	(0.174)
Born in 1960s	0.632***	0.459**	0.791***	0.309	-0.225	0.505
	(0.123)	(0.188)	(0.182)	(0.334)	(0.486)	(0.437)
Born in 1970s	0.771***	0.906***	0.627***	0.848*	0.464	0.870
	(0.176)	(0.273)	(0.221)	(0.452)	(0.475)	(0.831)
Born in 1980s	1.556***	2.002***	1.114***	2.374***	1.845***	2.538**
	(0.253)	(0.401)	(0.382)	(0.457)	(0.499)	(0.969)
Observations	16051	8098	7953	2457	1270	1187
Adjusted R^2	0.561	0.227	0.166	0.613	0.223	0.124

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples;

3) Unit: log of 10000 yuan; measured in 1990 price level

4) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation are included;

5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3-A10: Main estimation with brothers and sisters

	Rural sample			Urban sample		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Log Agricultural output p.c.- age 0 to 3	0.067	0.198	-0.107	-0.147**	0.010	-0.324**
	(0.171)	(0.249)	(0.245)	(0.064)	(0.086)	(0.135)
Log Industrial output p.c.- age 0 to 3	0.120	0.125	0.115	0.000	-0.130	0.111
	(0.083)	(0.103)	(0.109)	(0.112)	(0.188)	(0.091)
Log Gov. social expenditure p.c.- age 0 to 3	0.294***	0.388***	0.233*	0.403***	0.622***	0.084
	(0.099)	(0.140)	(0.123)	(0.129)	(0.221)	(0.224)
Death rate (‰) - age 0 to 3	-0.027**	-0.001	-0.051***	0.033	0.032	0.031
	(0.013)	(0.015)	(0.017)	(0.032)	(0.046)	(0.050)
Parental schooling years	0.063***	0.084***	0.040*	0.099***	0.138***	0.091**
	(0.016)	(0.024)	(0.022)	(0.031)	(0.038)	(0.039)
Father's occupation rank at child's age 14	0.002***	0.002**	0.002**	0.001	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Father's party membership at child's age 14	0.413***	0.364*	0.412**	0.016	-0.296	0.360
	(0.130)	(0.195)	(0.191)	(0.302)	(0.412)	(0.377)
Male	9.965***			11.273***		
	(0.095)			(0.240)		
Han Chinese	0.092	0.226	-0.021	-0.326	0.107	-1.060
	(0.260)	(0.368)	(0.338)	(0.811)	(0.805)	(1.058)
Number of brothers	-0.004	0.016	-0.017	0.213	0.344	0.074
	(0.047)	(0.063)	(0.065)	(0.141)	(0.219)	(0.214)
Number of sisters	0.067	0.146**	0.023	-0.200*	-0.078	-0.313*
	(0.058)	(0.070)	(0.075)	(0.107)	(0.238)	(0.161)
Birth order	0.076**	0.037	0.072	-0.100	-0.097	-0.086
	(0.037)	(0.054)	(0.048)	(0.110)	(0.188)	(0.155)
Exposure	2.069***	2.633***	1.246***	2.488***	2.795***	2.278***
	(0.290)	(0.459)	(0.429)	(0.403)	(0.756)	(0.758)
Exposure × Number of brothers	-0.551***	-0.590***	-0.265*	-0.643	-1.079*	0.016
	(0.111)	(0.159)	(0.136)	(0.473)	(0.597)	(0.681)
Exposure × Number of sisters	-0.351***	-0.327*	-0.327**	0.004	0.141	-0.448
	(0.117)	(0.168)	(0.139)	(0.235)	(0.407)	(0.403)
Observations	16051	8098	7953	2457	1270	1187
Adjusted R^2	0.562	0.229	0.165	0.614	0.226	0.127

Note:

1) Robust Standard Errors (SEs) are presented in parentheses;

2) SEs are clustered at the county and prefecture levels for the rural and urban samples;

3) Unit: log of 10000 yuan; measured in 1990 price level

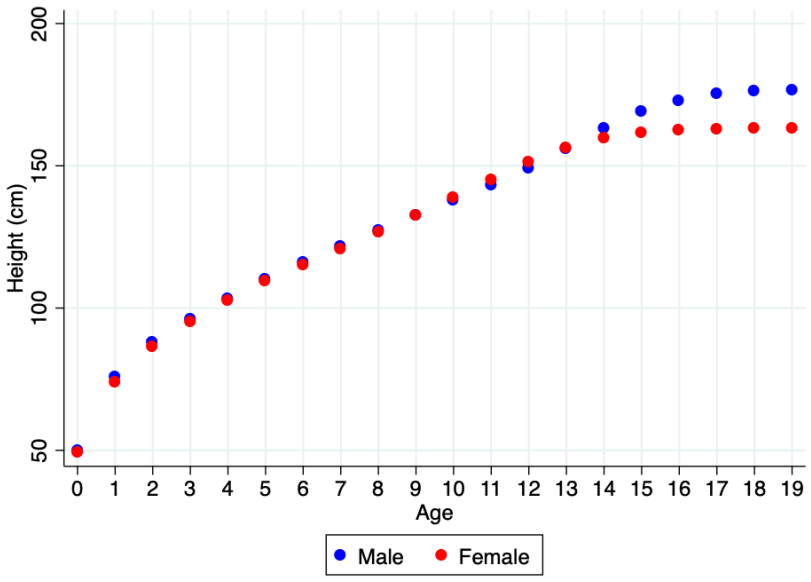
4) County fixed effects, a share of urban population in each county, and a group of dummy variables indicating the missing values for father's party membership and occupation as well as birth order variables are included;

5) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B: How the damping coefficients and the exposure variable are generated

The WHO Multicentre Growth Reference Study (MGRS) develops the growth references for infants and young children using longitudinal surveys. The sample used consists of about 8,500 children from six countries - Brazil, Ghana, India, Norway, Oman, and the United States [Onis et al., 2004].[1] Figure 3-B1 plots median height-for-age from the WHO Child Growth Standards. At age 0, the median height-for-age is about 50 cm for both females and males, and it jumps to 75 cm (female 74.005 cm; male 75.739 cm) at age 1, then further to about 87 cm (female 86.401 cm; male 87.802 cm) at age 2. As children growing up, the increment between two consecutive ages becomes smaller. At age 15, the growth for females are stop, whereas for males it stopped at around age 16. The height grows the most during the first two years of children's life.

Figure 3-B1: WHO Child Growth Standards: median height-for-age



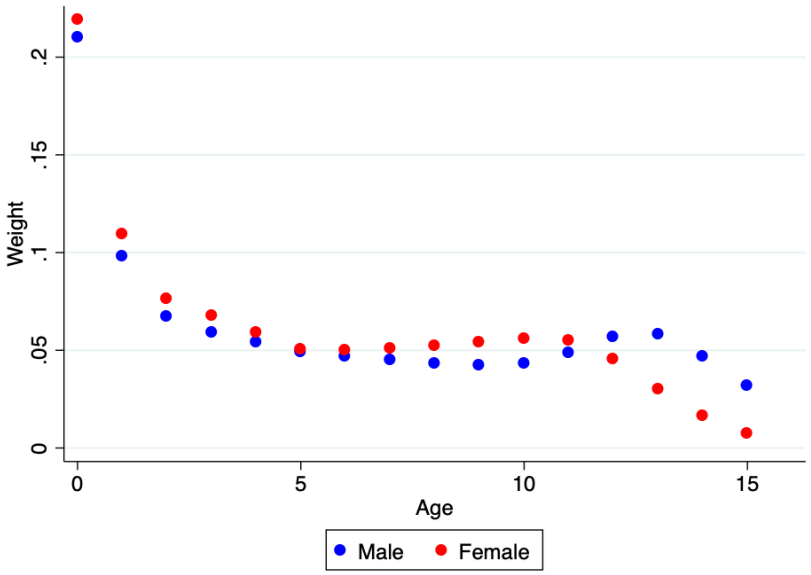
Using this growth reference data, we compute two variables: 1) the 'exposure' variable, which captures the share of an individual's height growing years (0-16) was at and after 1979, the year of the start of the economic reform and the introduction of the OCP and 2) the damping coefficient ω_n in equation (II).

To do so, we first calculate an increase in height between two consecutive ages as a proportion of the total increment between age 0 and 16. For example, female median height at age 0 is 49.15 cm and it increases to 74.01 cm at age 1. Female median height at age 16 is 162.516 cm. Thus the total increment in height between age 0 and 16 is 113.37 cm. The share of the first year increment to the total increment is 21.9% (=(74.01-49.15)/113.37).

[1]Data source: Boys 0-5 years old - https://www.who.int/childgrowth/standards/h_f_a_tables_z_boys/en/; Girls 0-5 years old - https://www.who.int/childgrowth/standards/h_f_a_tables_z_girls/en/; 5-19 years old - https://www.who.int/growthref/who2007_height_for_age/en/

The increment for each age is plotted in Figure 3-B2. The figure shows that between age 0 and 1 the height of an average child increases by more than 22% of its total height growth between age 0 and 16. Whereas between age 1 and 2 it increases by 11%, and between age 2 and 3 it increases by another 7.62%. Afterwards, the growth slows down significantly.

Figure 3-B2: Increments between consecutive ages

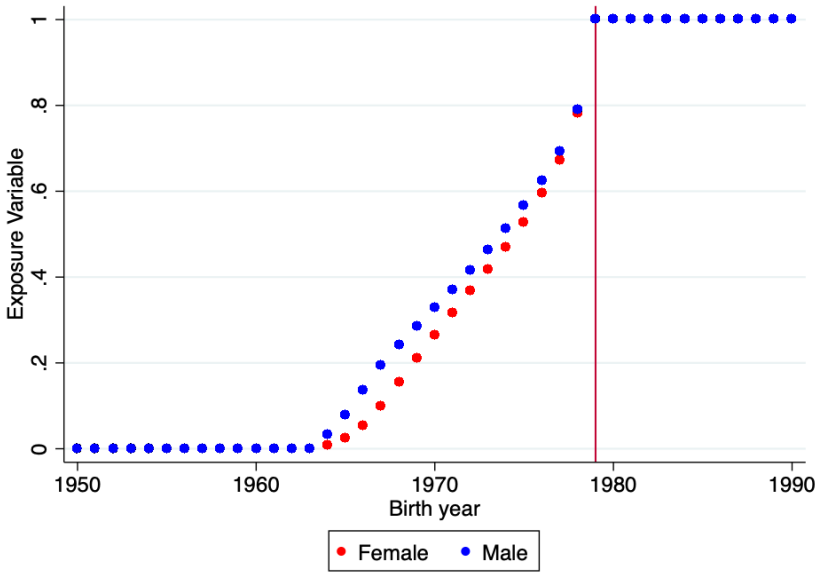


Second, to generate the 'exposure' variable, we assign individuals who were born before 1964 as zero and those who were born after 1979 as 1. In between, we use the cumulative share of the height increment in total increment depending on the individuals' age in 1979. For example, a female born in 1978, aged 1 in 1979, would have the exposure variable set equal to 1-0.219=0.781, and if she was born in 1977, the variable would be set to be equal to 0.671 (=1-0.219-0.11). Figure 3-B3 shows the distribution of the exposure variable for males (blue dots) and females (red dots) separately.

Third, we compute ω_n using the first three years (0-3) share of height increments in the total height growth between aged 0-16. In the above female example, we add up the three numbers and calculate the relative contribution of each age between age 0 and 3. These ωs are used to weight the local economic and social condition variables (E_n) to generate a weighted average value for these variables. ω_ns in equation (II) for females are calculated as:

$$\omega_n = 21.92594/(21.92594 + 10.9343 + 7.61502) \text{ for age 0-1}$$
$$\omega_n = 10.9343/(21.92594 + 10.9343 + 7.61502) \text{ for age 1-2}$$
$$\omega_n = 7.61502/(21.92594 + 10.9343 + 7.61502) \text{ for age 2-3}$$

Figure 3-B3: Cohort distribution of the 'exposure' variable



Appendix C: Local gazetteer data collection

C-1. Variable description

We collected annual data on the following variables from county gazetteers for the period between 1950 and 1990.

1) Industrial output value: This variable is the value of annual total industrial output produced in a county. Some counties use the corresponding year price, but others use a constant price across years. We convert all the values into the 1990 price using a provincial-level retail price index. The unit is ten thousand yuan.

2) Agricultural output value: This variable is the value of total agricultural output produced in a county. It includes farming, forestry, animal husbandry, and fishery. Some counties use the corresponding year price, but others use a constant price across years. We convert all the values into the 1990 price using a provincial-level retail price index. The unit is ten thousand yuan.

3) Local government expenditure on social services: This variable is the amount of annual local government expenditures on social services, including health care, education, social security benefit, culture, and science. We convert all the values into the 1990 price using a provincial-level retail price index. The unit is ten thousand yuan.

4) Death rate: The definition of death rate is the number of deaths divided by the total population for each year. The unit is ‰.

5) Rural *hukou* population: It is the number of people whose household registration (*hukou*) is agricultural in a county. We generate the rural and urban *hukou* share variables using this variable and county total population information.

6) Population: We generate per capita output and expenditure variables using total population data. The unit is persons.

C-2. Data source

Table 3-C1 provides data source. The main sources are local gazetteers and statistical yearbooks. China has changed its administrative divisions significantly over the last decades, so some regions have multiple gazetteers. We do not merge data from different sources unless available observations for the overlapping periods are the same across the books.

Table 3-C1: Data source

Province	Source
Anhui	1) Local gazetteers
Beijing	1) Local gazetteers 2) Comprehensive Statistical Data and Materials on 50 Years of New China
Chongqing	1) Local gazetteers
Fujian	1) Local gazetteers 2) Fujian Statistical Yearbook "New China 60 Years-Fujian"
Gansu	1) Local gazetteers 2) Gansu Population Statistical Yearbook 1949-1987; 3) Gansu Statistical Yearbook "New China 60 Years - Gansu"
Guangdong	1) Local gazetteers 2) 2 City Statistical Yearbooks
Guangxi	1) Local gazetteers 2) Guangxi Population Statistical Yearbook 1949-1985
Guizhou	1) Local gazetteers 2) Guizhou Population Statistical Yearbook 1949-1984 3) Guizhou Statistical Yearbook "Guizhou 60 Years 1949-2009" 4) Guizhou Statistical Yearbook "Guizhou 30 Year" 5) City Statistical Yearbook
Hebei	1) Local gazetteers 2) Hebei Statistical Yearbook "New Hebei 60 Years" 3) Hebei Statistical Yearbook "New Hebei 50 Years" 4) Hebei Province Population Statistical Yearbook 1949-1984
Heilongjiang	1) Local gazetteers 2) Heilongjiang Statistical Yearbook "Longjiang 60 Years" 3) Heilongjiang Population Statistical Yearbook
Henan	1) Local gazetteers 2) Henan Statistical Yearbook "Henan 30 Years"
Hubei	1) Local gazetteers 2) Hubei Statistical Yearbook 1985 and 1987 3) Hubei Statistical Yearbook "Hubei 30 Years"
Hunan	1) Local gazetteers 2) Hunan Population Statistical Yearbook 1949-1991 3) City statistical yearbook
Jiangsu	1) Local gazetteers 2) Jiangsu Population Statistical Yearbook 1949-1985; 3) Jiangsu Statistical Yearbook - "Jiangsu 60 Years"
Jiangxi	1) Local gazetteers 2) Jiangxi Statistical Yearbook "New China 50 Years - Jiangxi"; 3) Jiangxi Population Statistical Yearbook 1949-1985
Jilin	1) Local gazetteers 2) Jilin Population Statistical Yearbook 1949-1984
Liaoning	1) Local gazetteers 2) Liaoning Statistical Yearbook "Liaoning 60 Years" 3) Liaoning Statistical Yearbook "Liaoning 40 Years" 4) Liaoning Population yearbook 1949-1984 5) city statistical yearbook
Shaanxi	1) Local gazetteers
Shandong	1) Local gazetteers 2) Shandong Statistical Yearbook "New China 50 years - Shandong province" 3) Shandong Population Statistical Yearbook 1949-1984 4) City economic statistical yearbook
Shanghai	1) Local gazetteers 2) Shanghai Population Statistical Yearbook 1949-2000 3) Shanghai Statistical Yearbook - "Shanghai 60 Years of Statistics compilation"
Shanxi	1) Local gazetteers 2) Shanxi Population Statistical Yearbook 3) Shanxi Statistical Yearbook "Shanxi 60 Years"
Sichuan	1) Local gazetteers 2) Sichuan Statistical Yearbook "Sichuan Statistics Compilation (1979-1990)"
Tianjin	1) Local gazetteers 2) Tianjin Population Statistical Yearbook 1979-1988
Yunnan	1) Local gazetteers 2) Yunnan Population Statistical Yearbook 1949-1988
Zhejiang	1) Local gazetteers 2) Zhejiang Statistical Yearbook "Zhejiang 60 Years" 3) Zhejiang Population Statistical Yearbook 1949-1985
All provinces	China City Statistical Yearbook 1985, 1986, 1987, 1988, 1989, and 1990 Retail price index: Comprehensive Statistical Data and Materials on 50 Years of New China

C-3. Missing information

Table 3-C2 shows data source and the proportion of missing values for each variable. We collected 6492 agricultural output observations (by county and year) from the local gazetteers and 1988 observations from the statistical yearbooks. 47.2% of the total observations of agricultural output and 42.6% of industrial output observations are predicted owing to missing information. Local government social expenditure information are mostly from the local gazetteers and 25.0% of the total observations in our sample are predicted. About half of the collected mortality rate observations are from the gazetteers and the other half is mainly collected from provincial-level population statistical yearbooks. Its proportion of missing observation is the lowest among the four county-level variables.

Table 3-C2: Data source and the proportion of missing values

	Data source		**Missing Obs.**	**Total Obs.**	
	Gazetteer	Yearbook	(a)	(b)	(a/b)
Rural sample					
Agricultural output	6492	1988	7571	16051	0.472
Industrial output	7219	1990	6842	16051	0.426
Government expenditure	11758	258	4008	16051	0.250
Death rate	6889	6573	2589	16051	0.161
Urban sample					
Agricultural output	293	1359	805	2457	0.328
Industrial output	364	1214	879	2457	0.358
Government expenditure	248	1959	250	2457	0.102
Death rate	226	1731	500	2457	0.204

Column 1 shows the number of observations collected from local gazetteers, column 2 shows the number of observations collected from yearbooks, column 3 indicates the number of missing values.

C-4. Data prediction

To fill missing values, we impute them using available observations. The following equations (3-C1) and (3-C2) describe the main imputation methodologies.

$$\hat{x}_{c,p,t} = x_{c,t-1}\frac{x_{p,t}}{x_{p,t-1}} \tag{3-C1}$$

where $\hat{x}_{c,p,t}$ is a predicted value for a missing observation of county c in province p in year t. $x_{c,t-1}$ is a non-missing observation of the same county c in year $t-1$. The observations for province p are available for both years t and $t-1$. First, we compute the ratio between $x_{p,t}$ and $x_{p,t-1}$ at provincial level. Second, we multiply $x_{c,t-1}$ by this ratio and compute $\hat{x}_{c,p,t}$. This methodology requires an assumption that the growth rate of x between years t and $t-1$ at provincial level is the same as that of the county c. We use this method for all county-level variables included in equation (agricultural and industrial outputs, local government

expenditure on social services, and death rate).

There are cases that the gazetteers or yearbooks provide county-level gross domestic product (GDP) data, which covers the whole 40-year periods, while economic output information are not fully available. In this case, we impute missing values of the output variables using county-level GDP data as shown in equation 3-C2:

$$\hat{x}_{c,p,t} = gdp_{c,t}\frac{x_{p,t}}{gdp_{p,t}} \tag{3-C2}$$

where $\hat{x}_{c,p,t}$ is a predicted value of a missing observation of county c in provincial p at year t. Available observations are 1) $gdp_{c,t}$: GDP for county c in year t, 2) $x_{p,t}$: economic output for province p in year t, and 3) $gdp_{p,t}$: GDP for province p in year t. First, we compute the ration between GDP and economic output levels at provincial level. Then we multiply this ratio by the county-level GDP $gdp_{c,t}$ and calculate $\hat{x}_{c,p,t}$. This methodology requires an assumption that the ratio between GDP and economic output vales is the same between provincial-level and county-level data. More details on imputation methodologies are available upon request from the authors.

Appendix D: CHNS sample construction

The China Health and Nutrition Survey (CHNS) is a longitudinal survey of households and individuals conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention since 1989. We use all rounds between 1989 and 2011 (1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011), and the survey covers 9 provinces in China - Liaoning, Heilongjiang, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, and Guizhou.

There are 157,286 observations (including repeated individuals) between 1989 and 2011 waves, and 82,646 individuals of them were born between 1950 and 1990 and over 20 years old at the survey. Birth region information is a crucial part of our analysis, but the CHNS did not collect individual birth county information. Instead, it provides birth province of each household's head and her/his spouse but only 7,974 out of 82,646 observations have that information. We assume that all household members were born in the same province as the household head's, and we use the spouse birth province instead when the household head's birth province information is missing. As a result, 74,313 (19,151 individuals) out of 82,646 observations have birth province information. Among them 47,916 height observations are available for 15,149 individuals. To minimise the measurement error, we calculate the average height of available observations for each individual. We include 12,366 (81.6%) of 15,149 individuals who were born in the 9 CHNS survey provinces. The CHNS does not collect information on the *hukou* status at birth so we instead use the current *hukou* status to define the rural sample. This assumption is reasonable considering that 85% of individuals in the CFPS rural sample currently hold the rural *hukou* in 2010. The final CHNS rural sample consists of 7,048 individuals.

Appendix to Chapter 4

Table 4-A1: Middle school types in Korea

2005	Types	Entrants	%
Middle school	Total	703,800	100.00
	General	703,384	99.94
	Physical Education	245	0.03
	Arts	171	0.02

Source: Statistical Yearbook of Education 2005, Ministry of Education, Republic of Korea

Table 4-A2: High school types in Korea

	School types		Entrants	%
All			681,444	100.0
	General		492,615	72.3
	Special-purpose		17,673	2.6
		- Arts	5,742	0.8
		- Physical Education	1,289	0.2
		- Science	1,821	0.3
		- Foreign Language	8,821	1.3
	Vocational		171,156	25.1
		- Agricultural	5,450	0.8
		- Technical	63,151	9.3
		- Commercial	54,214	8.0
		- Fishery & Marine	1,443	0.2
		- Business	13,785	2.0
		- Comprehensive	33,113	4.9

Source: Statistical Yearbook of Education 2008, Ministry of Education, Republic of Korea

Table 4-A3: Number of schools per school district

	# School Districts		# Middle schools			# Middle schools per district		
Panel A: Middle school	EP (a)	non-EP (b)	Mixed (c)	All-girls (d)	All-boys (e)	(A) Mixed c/(a+b)	(B) All-girls d/(a+b)	(C) All-boys e/(a+b)
Metropolitan areas								
Seoul	46	0	292	44	48	6.3	1.0	1.0
Busan	23	0	101	33	8	4.4	1.4	0.3
Daegu	20	0	105	8	12	5.3	0.4	0.6
Incheon	25	1	78	26	30	3.0	1.0	1.2
Gwangju	13	0	75	7	8	5.8	0.5	0.6
Daejeon	32	4	68	9	11	1.9	0.3	0.3
Ulsan	24	0	55	4	4	2.3	0.2	0.2
Provinces								
Sejong Special Self-Governing City	10	1	17	1	1	1.5	0.1	0.1
Gyeonggi-do	230	30	557	31	31	2.1	0.1	0.1
Gangwon-do	108	4	117	21	24	1.0	0.2	0.2
Chungcheongbuk-do	72	28	96	16	16	1.0	0.2	0.2
Chungcheongnam-do	129	8	137	23	26	1.0	0.2	0.2
Jeollabuk-do	113	1	170	19	20	1.5	0.2	0.2
Jeollanam-do	190	4	198	25	27	1.0	0.1	0.1
Gyeongsangbuk-do	172	66	185	39	47	0.8	0.2	0.2
Gyeongsangnam-do	141	21	181	41	43	1.1	0.3	0.3
Jeju Special Self-Governing Province	28	2	31	7	7	1.0	0.2	0.2
	# School Districts		# High schools			# High schools per district		
Panel B: High school	EP (f)	non-EP (g)	Mixed (h)	All-girls (i)	All-boys (j)	Mixed c/(a+b)	All-girls d/(a+b)	All-boys e/(a+b)
Metropolitan areas								
Seoul	11	0	78	57	51	7.1	5.2	4.6
Busan	5	0	29	25	27	5.8	5.0	5.4
Daegu	2	0	22	12	14	11.0	6.0	7.0
Incheon	4	0	27	25	28	6.8	6.3	7.0
Gwangju	1	0	19	15	14	19.0	15.0	14.0
Daejeon	1	0	22	8	8	22.0	8.0	8.0
Ulsan	5	0	24	6	7	4.8	1.2	1.4
Provinces								
Sejong Special Self-Governing City	0	1	7	1	1	7.0	1.0	1.0
Gyeonggi-do	18	28	321	21	20	7.0	0.5	0.4
Gangwon-do	0	1	49	18	19	49.0	18.0	19.0
Chungcheongbuk-do	1	1	31	8	7	15.5	4.0	3.5
Chungcheongnam-do	1	1	41	18	13	20.5	9.0	6.5
Jeollabuk-do	3	1	41	28	25	10.3	7.0	6.3
Jeollanam-do	3	1	56	14	8	14.0	3.5	2.0
Gyeongsangbuk-do	1	1	51	35	33	25.5	17.5	16.5
Gyeongsangnam-do	4	1	86	29	23	17.2	5.8	4.6
Jeju Special Self-Governing Province	1	1	10	5	6	5.0	2.5	3.0

Source: https://schoolzone.edumac.kr/; Statistical yearbook of education 2016

Figure 4-A1: MSAT performance between excluded and included students

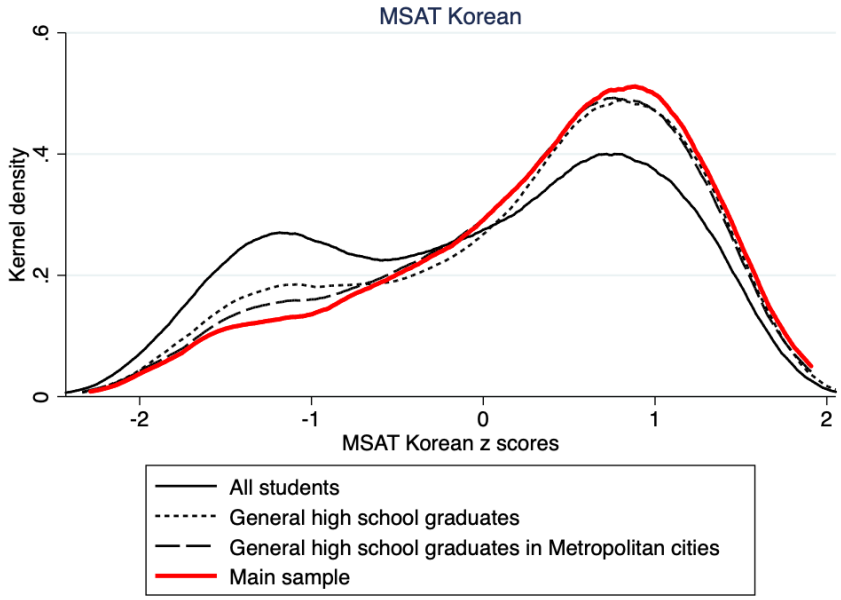


Table 4-A4: Middle school Female teacher ratio

	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
SSSS	-0.042	0.506	0.231	-0.867**	-0.612	-0.825*
	(0.356)	(0.308)	(0.403)	(0.431)	(0.402)	(0.439)
SSCO	-0.999	-0.019	-0.636	-1.350*	-1.264**	-1.534**
	(0.785)	(0.669)	(1.055)	(0.708)	(0.570)	(0.705)
COCO	-0.225	0.235	-0.427	-0.062	0.710	0.555
	(0.308)	(0.307)	(0.332)	(0.392)	(0.539)	(0.441)
SSSS × Middle school Female teacher ratio	0.470	-0.488	-0.020	1.313**	1.048*	1.066*
	(0.559)	(0.467)	(0.563)	(0.601)	(0.542)	(0.620)
SSCO × Middle school Female teacher ratio	1.711	0.330	1.181	1.934**	1.785**	2.186**
	(1.049)	(0.909)	(1.467)	(0.960)	(0.785)	(0.985)
COCO × Middle school Female teacher ratio	0.469	-0.398	0.724	0.126	-1.002	-0.742
	(0.444)	(0.431)	(0.463)	(0.541)	(0.731)	(0.620)
Middle school Female teacher ratio	0.100	0.730**	0.918**	-1.408*	-0.649	0.248
	(0.427)	(0.305)	(0.445)	(0.779)	(0.610)	(0.743)
Observations	636	636	636	716	716	716
Adjusted R^2	0.212	0.291	0.290	0.075	0.159	0.177

Note:
1) The dependent variables are the CSAT z-scores;
2) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4-A2: Two middle school group comparison - MSAT performance - (Left) Girls (Right) Boys

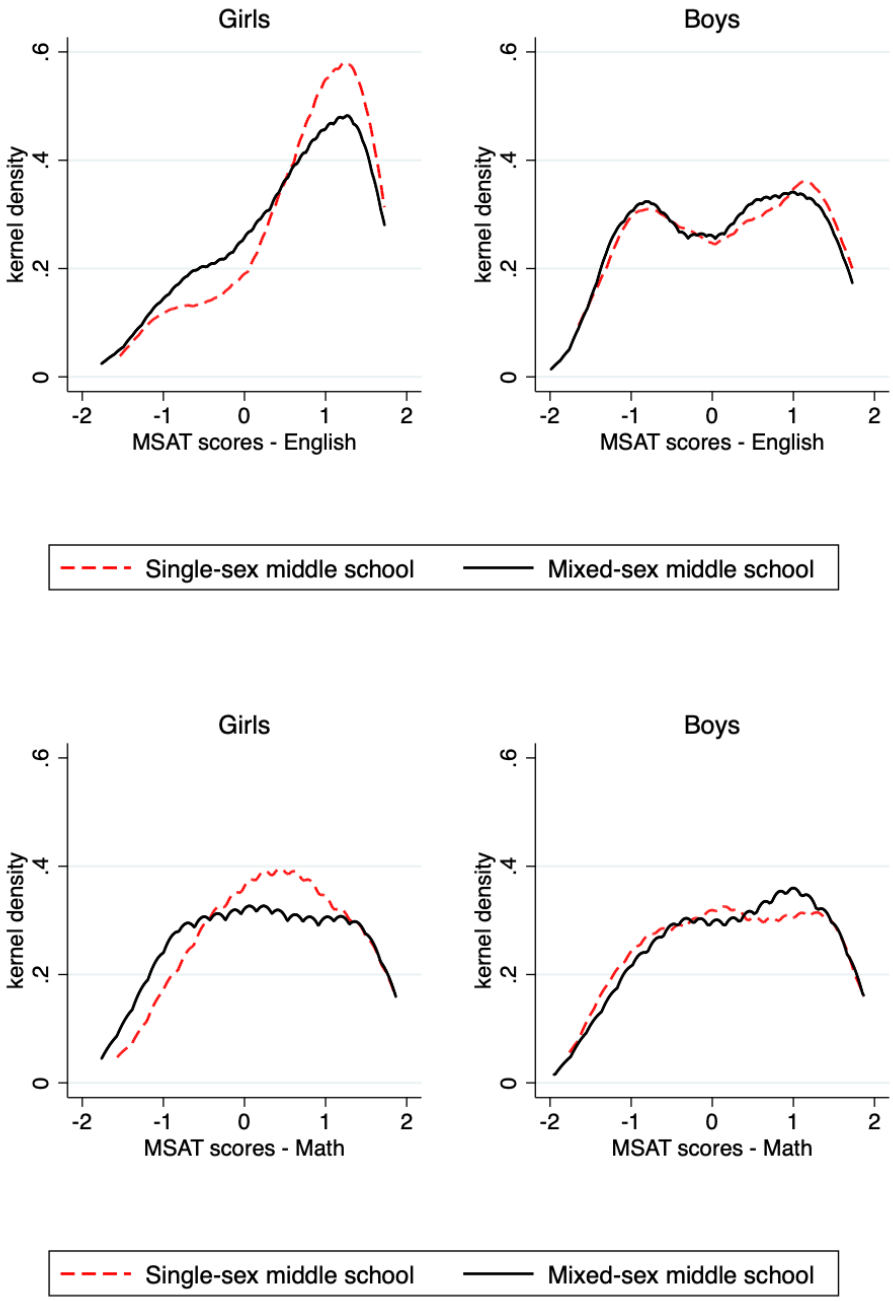


Table 4-A5: Peer interaction in class

	Middle school				High school			
Panel A: Girls	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SSSS	0.048	0.035	-0.061	-0.044	-0.041	-0.124**	-0.146**	-0.119**
	(0.062)	(0.048)	(0.053)	(0.061)	(0.051)	(0.057)	(0.067)	(0.045)
SSCO	-0.062	-0.094*	-0.173***	-0.090	-0.095	-0.207**	-0.140*	0.018
	(0.056)	(0.048)	(0.064)	(0.063)	(0.061)	(0.082)	(0.079)	(0.050)
COCO	-0.021	-0.024	-0.063	-0.043	0.005	-0.103*	-0.116*	-0.051
	(0.066)	(0.064)	(0.053)	(0.059)	(0.047)	(0.057)	(0.068)	(0.052)
Observations	628	628	628	628	628	628	628	628
R-squared	0.054	0.060	0.075	0.065	0.095	0.114	0.074	0.041
Mean Dep. var.	0.25	0.64	0.63	0.32	0.21	0.59	0.62	0.27

	Middle school				High school			
Panel B: Boys	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SSSS	0.062	0.023	0.014	-0.108**	-0.057	0.004	-0.090	-0.107**
	(0.050)	(0.047)	(0.059)	(0.047)	(0.064)	(0.040)	(0.055)	(0.043)
SSCO	0.029	0.068	0.091	-0.153*	0.030	0.064	0.076	-0.085
	(0.101)	(0.077)	(0.093)	(0.091)	(0.077)	(0.054)	(0.061)	(0.057)
COCO	0.035	0.007	0.024	0.017	-0.019	0.096*	-0.082	0.000
	(0.044)	(0.058)	(0.052)	(0.050)	(0.048)	(0.055)	(0.056)	(0.048)
Observations	697	697	697	697	697	697	697	697
R-squared	0.062	0.040	0.053	0.056	0.061	0.062	0.052	0.062
Mean Dep. var.	0.27	0.52	0.50	0.38	0.21	0.48	0.44	0.31

Dependent variables
(1) and (5) I help classmates who fall behind with school work - Yes=1, No=0;
(2) and (6) I discuss with my classmates to understand difficult class material - Yes=1, No=0;
(3) and (7) I share my class notes or material with my classmates - Yes=1, No=0;
(4) and (8) I enjoy working on group assignments - Yes=1, No=0;
Note:
1) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6;
2) Robust standard errors in parentheses are clustered by middle school;
3) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-A6: Public high school only

	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
SS-SS	-0.152	-0.322***	-0.236*	-0.092	-0.117	-0.035
	(0.135)	(0.107)	(0.123)	(0.170)	(0.167)	(0.166)
SS-CO	0.180	-0.323**	-0.122	0.233	0.027	-0.029
	(0.158)	(0.143)	(0.122)	(0.165)	(0.172)	(0.145)
CO-CO	0.106	-0.008	-0.087	0.170	0.218	-0.040
	(0.110)	(0.124)	(0.123)	(0.144)	(0.168)	(0.188)
Observations	316	316	316	268	268	268
Adjusted R^2	0.402	0.428	0.375	0.379	0.339	0.267

Note:
1) The dependent variables are the CSAT z-scores;
2) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-A7: Early admission application

	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
SS-SS	-0.091	-0.135*	-0.215***	-0.073	-0.108	-0.080
	(0.064)	(0.071)	(0.077)	(0.106)	(0.111)	(0.132)
SS-CO	0.039	-0.155	-0.340***	0.037	-0.021	-0.090
	(0.090)	(0.095)	(0.096)	(0.098)	(0.098)	(0.124)
CO-CO	-0.027	-0.048	-0.244**	0.096	0.009	-0.175
	(0.079)	(0.064)	(0.096)	(0.079)	(0.107)	(0.112)
Observations	624	624	624	703	703	703
Adjusted R^2	0.390	0.481	0.384	0.355	0.364	0.223

Note:
1) The dependent variables are the CSAT z-scores;
2) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6 in addition to two Soosi variables;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-A8: Including non-metropolitan areas

	CSAT - Female			CSAT - Male		
	(1)	(2)	(3)	(4)	(5)	(6)
SS-SS	-0.136**	-0.141**	-0.120*	0.034	0.012	-0.047
	(0.063)	(0.067)	(0.069)	(0.112)	(0.105)	(0.120)
SS-CO	0.019	-0.072	-0.360***	0.133	0.027	0.021
	(0.078)	(0.094)	(0.100)	(0.133)	(0.109)	(0.121)
CO-CO	-0.066	-0.135**	-0.286***	0.139	0.005	-0.188*
	(0.059)	(0.060)	(0.083)	(0.084)	(0.100)	(0.102)
Observations	888	875	865	936	919	942
Adjusted R^2	0.384	0.477	0.358	0.310	0.357	0.213

Note:
1) The same specification as Table 4-6 is used;
2) Robust standard errors in parentheses are clustered by middle school;
3) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$;

Figure 4-A3: Two high school group comparison - CSAT performance - (Left) Girls (Right) Boys

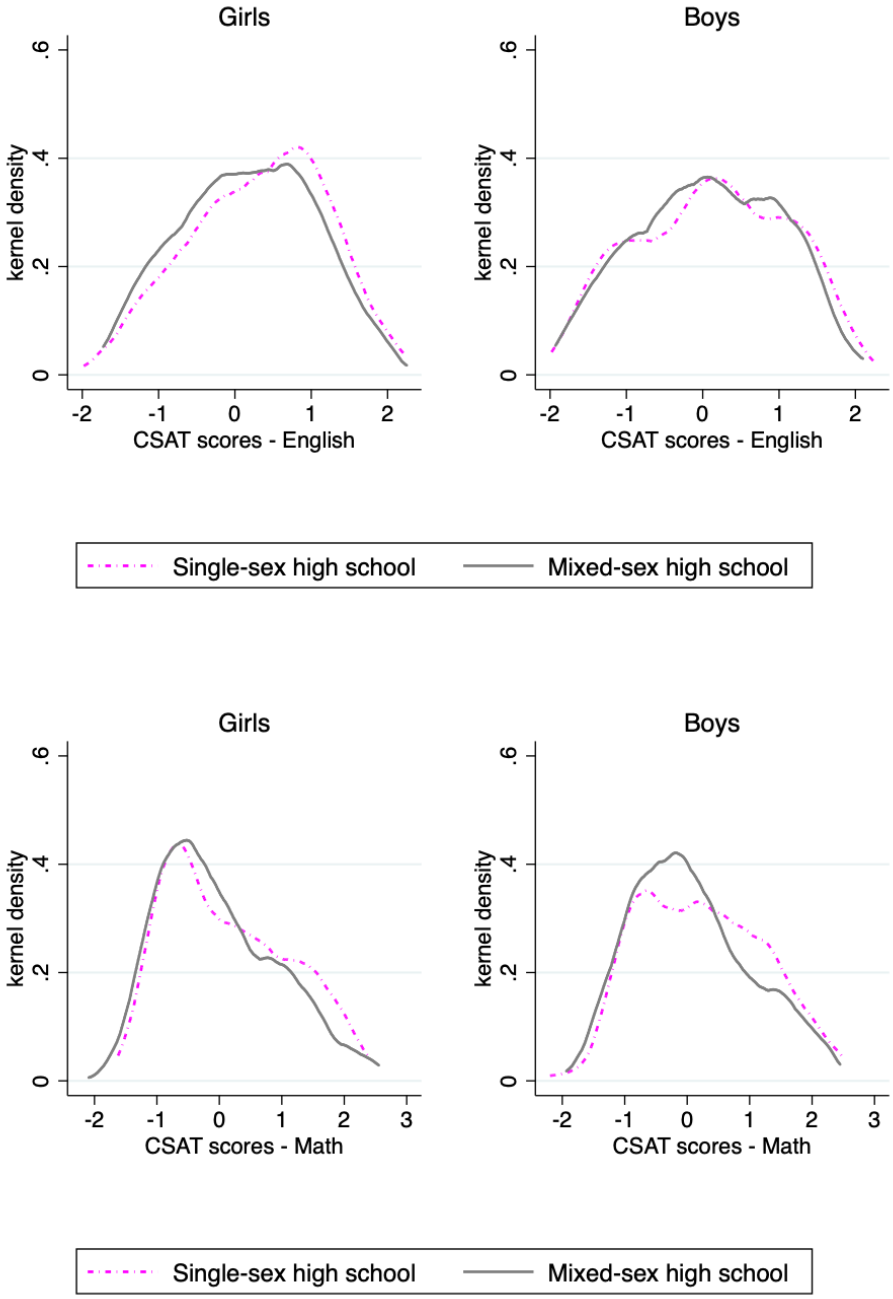


Figure 4-A4: Three group comparison - CSAT performance - (Left) Girls (Right) Boys

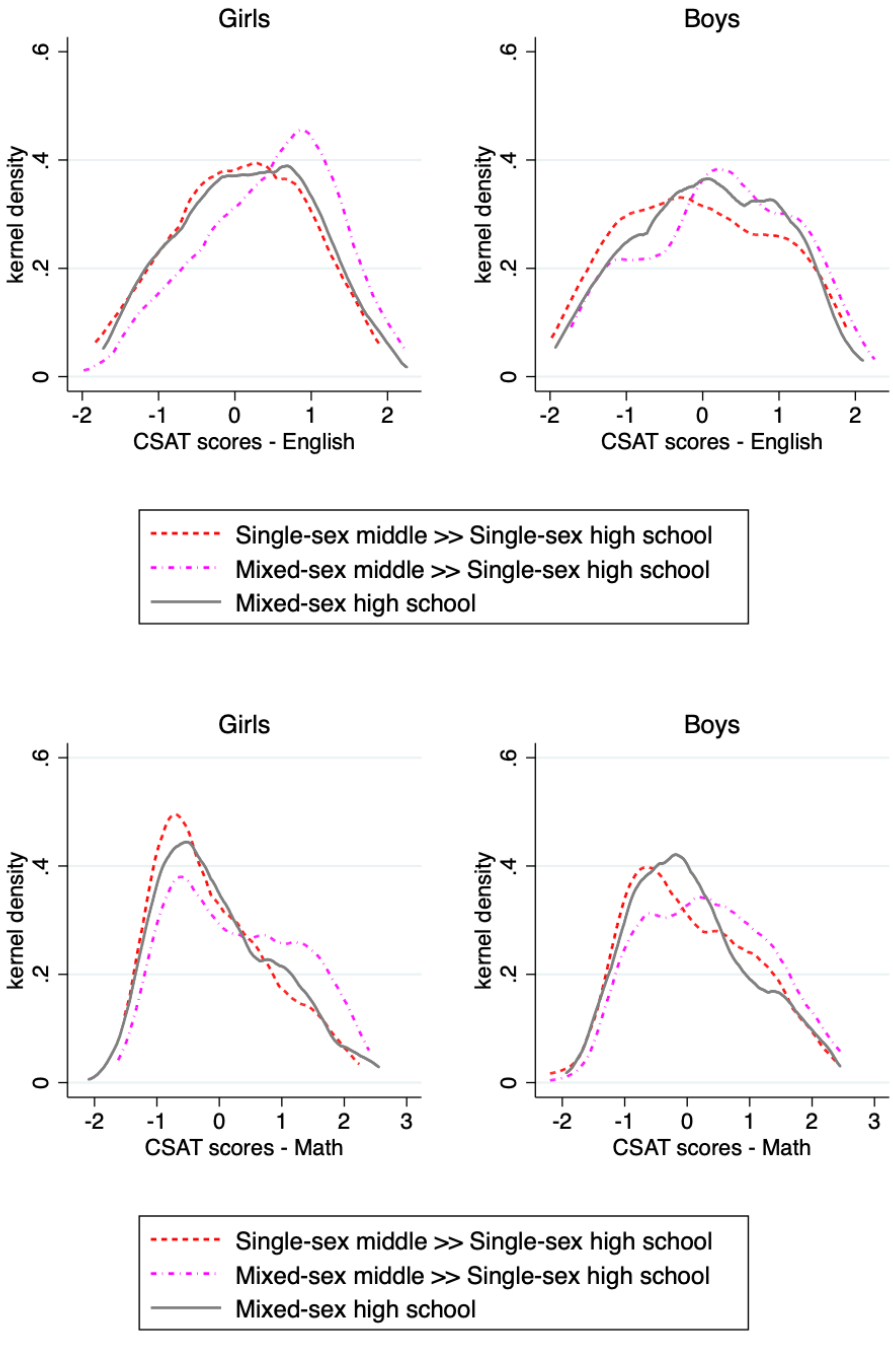


Figure 4-A5: Three group comparison - MSAT performance - (Left) Girls (Right) Boys

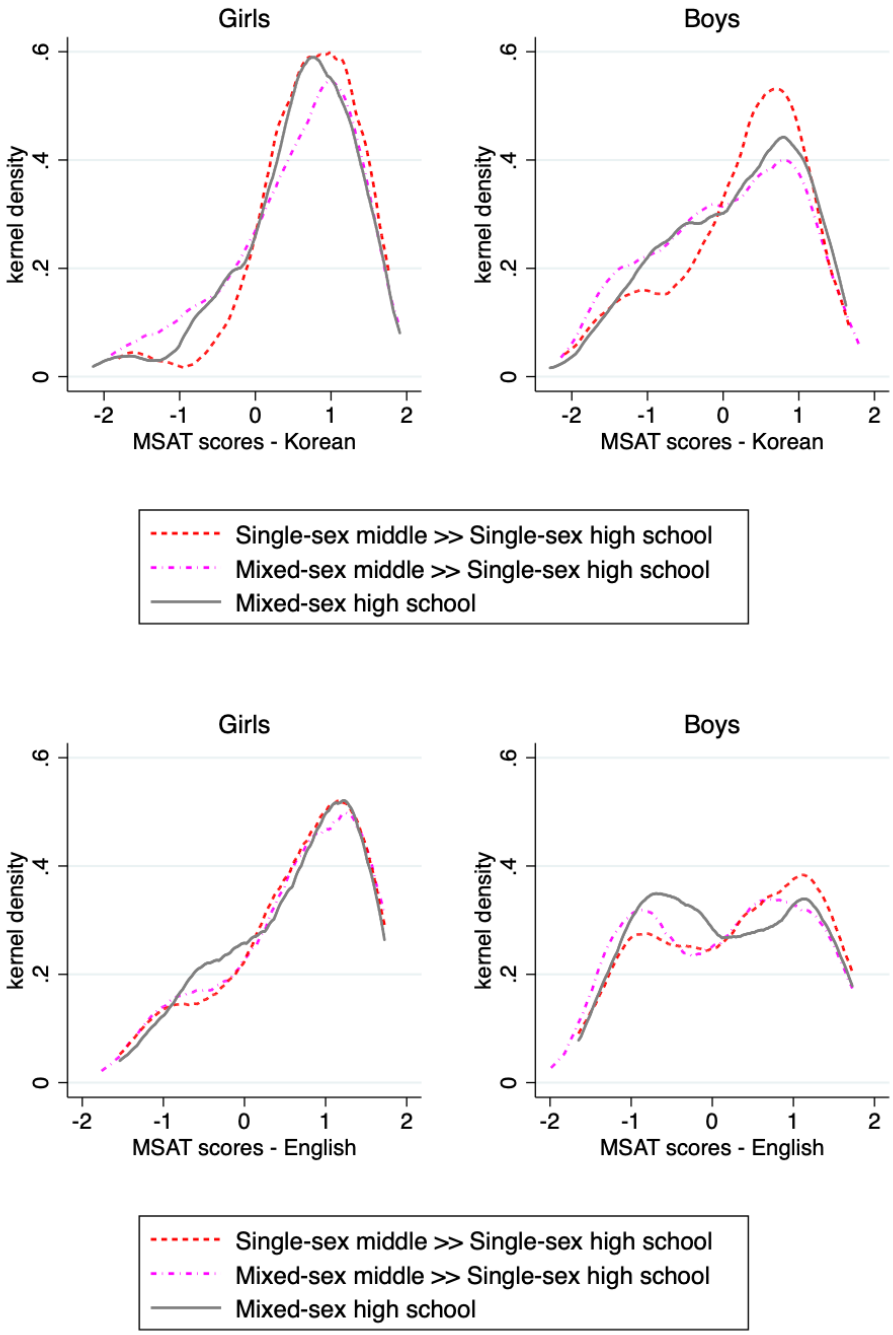


Figure 4-A6: The relationship between performance differences and high school female teacher ratio

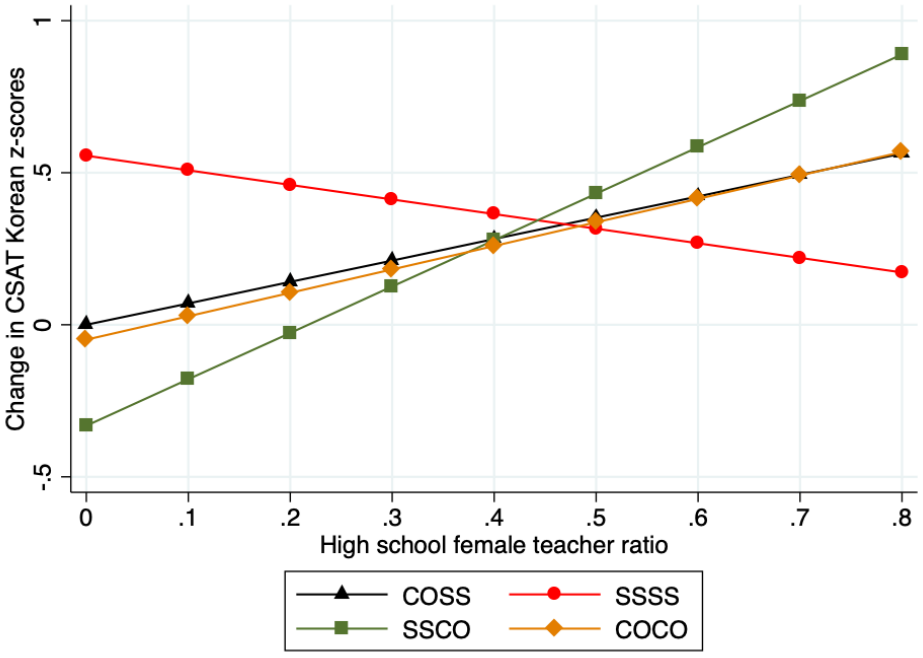


Table 4-A9: MSAT performance in the first three years - Girls

	1st Year			2nd Year			3rd Year		
	Kor (1)	Eng (2)	Math (3)	Kor (4)	Eng (5)	Math (6)	Kor (7)	Eng (8)	Math (9)
SSSS	0.242***	0.163	0.227**	0.131	0.138*	0.194**	0.278***	0.165**	0.217**
	(0.069)	(0.104)	(0.094)	(0.090)	(0.074)	(0.087)	(0.088)	(0.075)	(0.090)
SSCO	0.167**	0.105	0.221	0.090	0.086	0.205*	0.237***	0.245***	0.221**
	(0.069)	(0.123)	(0.135)	(0.088)	(0.087)	(0.104)	(0.088)	(0.072)	(0.106)
COCO	0.106	-0.030	0.071	0.063	0.003	0.056	0.116	-0.056	0.095
	(0.070)	(0.070)	(0.076)	(0.089)	(0.078)	(0.098)	(0.071)	(0.079)	(0.117)
Observations	628	634	630	621	625	627	636	636	636
Adjusted R^2	0.285	0.384	0.275	0.238	0.290	0.226	0.213	0.293	0.291

Note:
1) The dependent variables are the CSAT z-scores;
2) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-A10: MSAT performance in the first three years - Boys

	1st Year			2nd Year			3rd Year		
	Kor (1)	Eng (2)	Math (3)	Kor (4)	Eng (5)	Math (6)	Kor (7)	Eng (8)	Math (9)
SSSS	-0.237**	-0.083	-0.085	-0.094	-0.205*	-0.029	0.051	0.117	-0.080
	(0.108)	(0.139)	(0.106)	(0.096)	(0.110)	(0.100)	(0.122)	(0.109)	(0.107)
SSCO	-0.216	-0.081	-0.108	-0.028	-0.204*	0.132	-0.023	-0.059	-0.047
	(0.146)	(0.130)	(0.142)	(0.133)	(0.116)	(0.127)	(0.111)	(0.124)	(0.137)
COCO	0.018	-0.001	-0.050	0.138	-0.014	0.036	0.043	0.012	0.040
	(0.074)	(0.101)	(0.111)	(0.103)	(0.112)	(0.129)	(0.087)	(0.098)	(0.118)
Observations	707	712	711	701	703	704	716	716	716
Adjusted R^2	0.165	0.254	0.186	0.082	0.163	0.118	0.070	0.151	0.170

Note:
1) The dependent variables are the CSAT z-scores;
2) Column (1)-(6) include all the covariates in column (4)-(6) in Table 4-6;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-A11: Main results using the raw test scores - Girls

	MSAT			CSAT		
	Korean	English	Math	Korean	English	Math
	(1)	(2)	(3)	(4)	(5)	(6)
SSSS	5.822***	4.401**	5.685**	-0.106*	-0.121*	-0.201***
	(1.855)	(1.998)	(2.364)	(0.061)	(0.065)	(0.068)
SSCO	4.969***	6.540***	5.802**	0.037	-0.151*	-0.329***
	(1.844)	(1.927)	(2.767)	(0.085)	(0.089)	(0.090)
COCO	2.436	-1.480	2.492	-0.023	-0.060	-0.239**
	(1.482)	(2.106)	(3.053)	(0.073)	(0.062)	(0.090)
Observations	636	636	636	636	636	636
Adjusted R^2	0.213	0.293	0.291	0.389	0.483	0.382

Note:
1) The dependent variables in column 1-3 are the MSAT raw scores, ranging 0 to 100, and in column 4-6 are the CSAT raw scores (the standardised scores given to students from the government), ranging -3.65 to 2.65 in the raw data;
2) The same specification as Table 4-6 is used;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4-A12: Main results using the raw test scores - Boys

	MSAT			CSAT		
	Korean	English	Math	Korean	English	Math
	(1)	(2)	(3)	(4)	(5)	(6)
SSSS	1.062	3.120	-2.094	-0.076	-0.100	-0.079
	(2.568)	(2.897)	(2.806)	(0.104)	(0.106)	(0.122)
SSCO	-0.479	-1.576	-1.242	0.057	-0.009	-0.090
	(2.336)	(3.305)	(3.591)	(0.096)	(0.098)	(0.117)
COCO	0.911	0.329	1.054	0.094	-0.006	-0.182*
	(1.817)	(2.616)	(3.097)	(0.079)	(0.108)	(0.107)
Observations	716	716	716	716	716	716
Adjusted R^2	0.070	0.151	0.170	0.356	0.362	0.221

Note:
1) The dependent variables in column 1-3 are the MSAT raw scores, ranging 0 to 100, and in column 4-6 are the CSAT raw scores (the standardised scores given to students from the government), ranging -3.65 to 2.65 in the raw data;
2) The same specification as Table 4-6 is used;
3) Robust standard errors in parentheses are clustered by middle school;
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4-A7: Geographic coverage of the main sample

Bibliography

Encyclopedia of Korean Culture. Academy of Korean Studies, 2000. (cited on pages 78 and 79)

Statistical Yearbook of Education. Department of Education, Korean Educational Development Institute, 2005. (cited on page 78)

Statistical Yearbook of Education. Department of Education, Korean Educational Development Institute, 2008. (cited on page 78)

D. Almond, H. Li, and S. Zhang. Land reform and sex selection in China. *Journal of Political Economy*, 127(2), 2019. (cited on pages 17 and 51)

J. Angrist, V. Lavy, and A. Schlosser. Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4), 2010. (cited on page 23)

J. D. Angrist and W. N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3), 1998. (cited on page 16)

J. D. Angrist, V. Lavy, and A. Schlosser. New evidence on the causal link between the quantity and quality of children. *NBER Working Paper*, 11835, 2005. (cited on page 23)

B. Ashton, K. Hill, A. Piazza, and R. Zeitz. Famine in China, 1958-61. *Population and Development Review*, 10(4):613–645, 1984. (cited on page 43)

L. Babcock, M. P. Recalde, L. Vesterlund, and L. Weingart. Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3), 2017. (cited on pages 76 and 98)

R. E. Bailey, T. J. Hatton, and K. Inwood. Health, height, and the household at the turn of the twentieth century. *Economic History Review*, 69(1):35–53, 2016. (cited on pages 39 and 60)

R. E. Bailey, T. J. Hatton, and K. Inwood. Atmospheric pollution, health, and height in late nineteenth century Britain. *The Journal of Economic History*, 78(4):1210–1247, 2018. (cited on page 39)

J. Bainster. *China's Changing Population*. Standford University Press, 1987. (cited on page 10)

S. H. Barcellos, L. S. Carvalho, and A. Lleras-Muney. Child gender and parental investments in India: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, 6(1), 2014. (cited on page 7)

J. Baten and J. E. Murray. Heights of men and women in 19th-century Bavaria: Economic, nutritional, and disease influences. *Explorations in Economic History*, 37:351–369, 2000. (cited on page 39)

A. S. Beard and M. J. Blaser. The ecology of height: The effect of microbial transmission on human height. *Perspectives in Biology and Medicine*, 45(4), 2002. (cited on page 39)

G. S. Becker and H. G. Lewis. On the interaction between the quantity and quality of children. *Journal of Political Economy*, 81(2), 1973. (cited on pages 1, 3, 31, 57, and 62)

C. Bettelheim. Economic reform in China. *Journal of Development Studies*, 1988. (cited on page 42)

Z.-W. Bi and C.-Y. Ji. Secular growth changes in body height and weight in children and adolescents in Shandong, China between 1939 and 2000. *Annals of Human Biology*, 32(5): 650–665, 2005. (cited on page 40)

R. S. Bigler and L. S. Liben. Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16(3): 162–166, 2007. (cited on pages 75 and 95)

S. E. Black, P. J. Devereux, and K. G. Salvanes. The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2), 2005. (cited on pages 17, 23, and 36)

G. Bloom and G. Xingyuan. Health sector reform: Lessons from China. *Social Science & Medicine*, 45(3):351–360, 1997. (cited on page 46)

D. Blumenthal and W. Hsiao. Privatization and its discontents - the evolving chinese health care system. *The New England Journal of Medicine*, 353(11):1165–1170, 2005. (cited on page 46)

S. Bo. Son preference, children's gender and parents' time allocation: evidence from China. *Applied Economics*, 50(45), 2018. (cited on page 7)

A. Booth and P. Nolen. Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81:542–555, 2012a. (cited on pages 75, 76, and 98)

A. Booth and P. Nolen. Gender differences in risk behaviour: Does nurture matter? *The Economic Journal*, 112, 2012b. (cited on pages 75 and 76)

A. Booth, L. Cardona-Sosa, and P. Nolen. Gender differences in risk aversion: Do single-sex environments affect their development? *Journal of Economic Behavior & Organization*, 99: 126–154, 2014. (cited on pages 75, 76, and 98)

A. Booth, X. Meng, E. Fan, and D. Zhang. The intergenerational behavioural consequences of a socio-political upheaval. *IZA Discussion Paper Series*, 11991, 2018a. (cited on pages 44 and 76)

A. Booth, E. Fan, X. Meng, and D. Zhang. Gender differences in willingness to compete: The role of culture and institutions. *The Economic Journal*, 129, 2019. (cited on page 16)

A. L. Booth, L. Cardona-Sosa, and P. Nolen. Do single-sex classes affect academic achievement? an experiment in a coeducational university. *Journal of Public Economics*, 168:109–126, 2018b. (cited on pages 75, 76, 77, 90, 92, and 94)

E. Borensztein and J. D. Ostry. Accounting for China's growth performance. *The American Economic Review, Papers and Proceedings*, 86(2):224–228, 1996. (cited on page 44)

V. K. Borooah. Gender bias among children in India in their diet and immunisation against disease. *Social Science & Medicine*, 58(9), 2004. (cited on page 7)

Y. Cai. China's below-replacement fertility: Government policy or socioeconomic development? *Population and Development Review*, 36(3):419–440, September 2010. (cited on pages 8 and 45)

L. Cameron and X. Meng. China's one child policy. *The New Palgrave Dictionary of Economics*, 2014. (cited on pages 8, 20, and 125)

L. Cameron, N. Erkal, L. Gangadharan, and X. Meng. Little emperors: Behavioral impacts of China's one-child policy. *Science*, 339(22), 2013. (cited on page 5)

S. E. Carrell, M. E. Page, and J. E. West. Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics*, 125(3), 2010. (cited on page 96)

C. Center for Population Studies. The major events of China's population activities. In *The China's Population Yearbook (1985)*, pages 1263–1288. Chinese Social Sciences Press, Beijing, 1986. (cited on pages 8, 10, 45, and 47)

Q. Chen. Interrupted maternal education and child health: The long run health impact of the chinese cultural revolution. Job Market Paper, 2010. (cited on page 40)

Y. Chen and L.-A. Zhou. The long-term health and economic consequences of the 1959-1961 famine in China. *Journal of Health Economics*, 26:659–681, 2007. (cited on pages 40 and 43)

E. J. Choi, H. R. Moon, and G. Ridder. Estimation of an education production function under random assignment with selection. *American Economic Review: Papers & Proceedings*, 104(5): 206–211, 2014. (cited on pages 75 and 76)

G. C. Chow. Capital formation and economic growth in China. *The Quarterly Journal of Economics*, 108(3):809–842, August 1993. (cited on page 43)

A. J. Coale. Excess female mortality and the balance of the sexes in the population: An estimate of the number of "missing females". *Population and Development Review*, 17(3), 1991. (cited on page 7)

W. H. Colby, F. W. Crook, and S.-E. H. Webb. Agricultural statistics of the People's Republic of China, 1949-90. *Statistical Bulletin 154783, United States Department of Agriculture*, 1992. (cited on page 46)

D. Conley and R. Glauber. Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of Human Resources*, 2006. (cited on page 23)

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 2019. (cited on page 7)

M. Das Gupta, J. Zhenghua, L. Bohua, X. Zhenming, W. Chung, and B.-H. Ok. Why is son preference so persistent in east and South Asia? a cross-country study of China, India and the republic of korea. *Journal of Development Studies*, 2003. (cited on pages 3 and 7)

Z. Dong and M. R. Phillips. Evolution of China's health-care system. *The Lancet*, 372, November 2008. (cited on page 44)

C. Dustmann, H. Ku, and D. W. Kwak. Why are single-sex schools successful? *Labour Economics*, 54:79–99, 2018. (cited on pages 75, 77, 80, 90, 91, 92, and 94)

A. Ebenstein. The "missing girls" of China and the unintended consequences of the one child policy. *Journal of Human Resources*, 45(1):87–115, 2010. (cited on pages 8, 17, 20, 60, and 125)

A. Ebenstein and S. Leung. Son preference and access to social insurance: Evidence from China's rural pension program. *Population and Development Review*, 36(1):47–70, 2010. (cited on pages 3 and 7)

L. Edlund, H. Li, J. Yi, and J. Zhang. Sex ratios and crime: Evidence from China's one-child policy. *IZA Discussion Paper Series*, 3214, 2007. (cited on page 47)

G. Eisenkopf, Z. Hessami, U. Fischbacher, and H. W. Ursprung. Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. *Journal of Economic Behavior & Organization*, 115:123–143, 2015. (cited on pages 75, 77, 90, 92, and 94)

R. A. Fabes, E. Pahlke, A. Z. Borders, and K. Galligan. US principals' attitudes about and experiences with single-sex schooling. *Educational Studies*, 41(3):293–311, 2015. (cited on pages 75 and 95)

E. Fan, X. Meng, Z. Wei, and G. Zhao. Rates of return to four-year university education: An application of regression discontinuity design. *The Scandinavian Journal of Economics*, 145: 249–260, 2018. (cited on page 36)

H. Fang, K. N. Eggleston, J. A. Rizzo, S. Rozelle, and R. J. Zeckhauser. The returns to education in China: Evidence from the 1986 compulsory education law. *NBER Working Paper No. 18189*, 2012. (cited on page 71)

G. Feeney and F. Wang. Parity progression and birth intervals in China: The influence of policy in hastening fertility decline. *Population and Development Review*, 19(1):61–101, 1993. (cited on pages 8 and 45)

A. Fernihough and M. E. McGovern. Physical stature decline and the health status of the elderly population in England. *Economics and Human Biology*, 16:30–44, 2015. (cited on pages 51, 53, and 56)

M. Francesconi and J. J. Heckman. Child development and parental investment: Introduction. *The Economic Journal*, 126, 2016. (cited on page 7)

J. L. Garcia. Fertility after China's (more-than) one-child policy. *John E. Walker Department of Economics, Clemson University, working paper*, 2019. (cited on pages 4 and 10)

A. Goldstein and S. Goldstein. China's labor force: The role of gender and residence. *Journal of Women and Gender Studies*, 1:87–118, 1990. (cited on page 11)

S. Goldstein. orms of mobility and their policy implications: Thailand and China compared. *Social Forces*, 65, 1987. (cited on page 11)

Y.-L. Gong and L.-M. Chao. The role of barefoot doctors. *American Journal of Public Health*, 72 (Supplement), 1982. (cited on page 45)

T. Gørgens, X. Meng, and R. Vaithianathan. Stunting and selection effects of famine: A case study of the Great Chinese Famine. *Journal of Development Economics*, 97:99–111, 2012. (cited on pages 40, 43, 48, and 51)

B. Gu, W. Feng, G. Zhigang, and Z. Erli. China's local national fertility policies at the end of the twentieth century. *Population and Development Review*, 33(1):129–147, 2007. (cited on pages 20 and 122)

R. Guo, H. Li, J. Yi, and J. Zhang. Fertility, household structure, and parental labor supply: Evidence from China. *Journal of Comparative Economics*, 46:145–156, 2018. (cited on page 5)

d. M. Haan. Birth order, family size and educational attainment. *Economics of Education Review*, 29(4), 2010. (cited on page 23)

Y. Hahn, L. C. Wang, and H.-S. Yang. Does greater school autonomy make a difference? evidence from a randomized natural experiment in South Korea. *Journal of Public Economics*, 2018. (cited on pages 79, 87, and 99)

D. F. Halpern, L. Eliot, R. S. Bigler, R. A. Fabes, L. D. Hanish, J. Hyde, L. S. Liben, and C. L. Martin. The pseudoscience of single-sex schooling. *Science*, 333, 2011. (cited on pages 75 and 95)

T. J. Hatton. How have Europeans grown so tall? how have Europeans grown so tall? how have Europeans grown so tall? *Oxford Economic Papers*, 66:349–372, 2014. (cited on pages 39 and 60)

T. J. Hatton. Stature and sibship: Historical evidence. *Centre for Economic History, The Australian National University Discussion Paper Series*, 2015-09, 2015. (cited on page 60)

T. J. Hatton and R. M. Martin. Fertility decline and the heights of children in Britain, 1886-1938. *Explorations in Economic History*, 47(4):505–519, October 2010. (cited on page 60)

T. Hesketh and X. Z. Wei. Health in China: From Mao to market reform. *BMJ*, 24(314), 1997. (cited on pages 44 and 45)

L. J. Hilliard and L. S. Liben. Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child Development*, 81(6):1787–1798, December 2010. (cited on pages 76 and 95)

D. Hipgrave. Communicable disease control in China: From mao to now. *Journal of Global Health*, 1(2), December 2011. (cited on pages 44 and 45)

A. Hu. *The Cultural Revolution (1966-1976)*. Enrich Professional Publishing (S) Private, Limited, 2013. (cited on page 45)

W. Huang, X. Lei, G. Ridd, J. Strauss, and Y. Zhao. Health, height, height shrinkage, and SES at older ages: Evidence from China. *American Economic Journal: Applied Economics*, 5(2): 86–121, 2013. (cited on pages 51 and 56)

W. Huang, X. Lei, and A. Sun. The great expectations: Impact of one-child policy on education of girls. *IZA Discussion Paper*, 9301, 2015. (cited on page 17)

Y. Huang. The sick man of Asia - China's health crisis. *Foreign Affairs*, 90(6):119–136, 2011. (cited on page 46)

J. N. Igbo, V. C. Onu, and O. N. O. Impact of gender stereotype on secondary school students' self-concept and academic achievement. *SAGE Open*, pages 1–10, January-March 2015. (cited on pages 76 and 95)

C. K. Jackson. Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics*, 96: 173–187, 2012. (cited on page 75)

C. K. Jackson. The effect of single-sex education on test scores, school completion, arrests, and teen motherhood: Evidence from school transitions. *NBER Working Paper*, 22222, 2016. (cited on pages 75, 90, 92, and 94)

S. Jayachandran. The roots of gender inequality in developing countries. *Annual Review of Economics*, 7, 2015. (cited on pages 3, 4, 7, and 30)

S. Jayachandran and I. Kuziemko. Why do mothers breastfeed girls less than boys? evidence and implications for child health in India. *Quarterly Journal of Economics*, 126, 2005. (cited on page 7)

C. Kang, C. Park, and M.-J. Lee. Effects of ability mixing in high school on adulthood earnings: quasiexperimental evidence from South Korea. *Journal of Population Economics*, 20:269–297, 2007. (cited on page 79)

S. Klasen. "Missing women" reconsidered. *World Development*, 22(7), 1994. (cited on page 7)

S. Klasen and C. Wink. "Missing women": Revisiting the debate. *Feminist Economics*, 9(2-3), 2003. (cited on page 7)

M. Kubo and A. Chaudhuri. Gender gap in health status of children in the context of one-child policy in China: Is it sibling rivalry or son preference? *Journal of Family and Economic Issues*, 38:204–217, 2017. (cited on pages 6, 40, 60, and 62)

F. Lai. Are boys left behind? the evolution of the gender achievement gap in Beijing's middle schools. *Economics of Education Review*, 29:383–399, 2010. (cited on page 7)

J. Lee. Sibling size and investment in children's education: an Asian instrument. *Journal of Population Economics*, 21(4), 2008. (cited on pages 17 and 23)

S. Lee, M. Niederle, and N. Kang. Do single-sex schools make girls more competitive? *Economics Letters*, 124, 2014. (cited on pages 75, 79, and 85)

S. Lee, L. J. Turner, S. Woo, and K. Kim. The impact of school and classroom gender composition on educational achievement. Working Paper, 2016. (cited on pages 76, 79, 80, 91, 96, and 99)

V. E. Lee and A. S. Bryk. Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology*, 78(5):381–395, 1986. (cited on pages 77 and 94)

V. E. Lee and M. E. Lockheed. The effects of single-sex schooling on achievement and attitudes in Nigeria. *Comparative Education Review*, 34(2), 1990. (cited on pages 75 and 94)

S. T. Lereya, S. Eryigit-Madzwamuse, C. Patra, J. H. Smith, and D. Woke. Body-esteem of pupils who attended single-sex versus mixed-sex schools: A cross-sectional study of intra-sexual competition and peer victimization. *Journal of Adolescene*, 37:1109–1119, 2014. (cited on page 98)

B. Li and H. Zhang. Does population control lead to better child quality? evidence from China's one-child policy enforcement. *Journal of Comparative Economics*, 45, 2017. (cited on page 28)

H. Li, J. Zhang, and Y. Zhu. The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*, 45(1), 2008. (cited on pages 5, 6, 23, and 28)

H. Li, J. Yi, and J. Zhang. Estimating the effect of the one-child policy on the sex ratio imbalance in China: Identification based on the difference-in-differences. *Demography*, 48, 2011. (cited on pages 5, 7, 10, and 20)

J. Li and R. S. Cooney. Son preference and the one child policy in China: 1979-1988. *Population Research and Policy Review*, 12(3), 1993. (cited on page 11)

Q. Li and L. An. Intergenerational health consequences of the 1959–1961 Great Famine on children in rural China. *Economics and Human Biology*, 18:27–40, 2015. (cited on page 40)

W. Li and D. T. Yang. The great leap forward: Anatomy of a central planning disaster. *Journal of Political Economy*, 2005. (cited on page 43)

M. H. Liang, P. S. Eichling, L. J. Fine, and G. J. Annas. Chinese health care: Determinants of the system. *Am J Public Health*, 63(2), 1973. (cited on page 42)

Y. Liang and Z. Dong. Has education led to secularization? based on the study of compulsory education law in China. *China Economic Review*, 54:324–336, 2019. (cited on page 71)

J. Lim and J. Meer. The impact of teacher–student gender matches: Random assignment evidence from South Korea. *Journal of Human Resources*, 52(4), 2017. (cited on page 96)

J. Y. Lin. The household responsibility system in China's rural reform. In *Proceedings of the Twentieth International Conference of Agricultural Economists*, 1988. (cited on pages 43 and 45)

J. Y. Lin. Collectivization and US's agricultural crisis in 1959-1961. *Journal of Political Economy*, 98(6):1228–1252, 1990. (cited on pages 43 and 44)

J. Y. Lin. Rural reforms and agricultural growth in China. *The American Economic Review*, 82 (1):34–51, March 1992. (cited on page 45)

H. Liu. The quality-quantity trade-off: evidence from the relaxation of China's one-child policy. *Journal of Population Economics*, 27:565–602, 2014. (cited on pages 5, 6, 28, 40, 60, and 62)

Y. Liu. Development of the rural health insurance system in China. *Health Policy and Planning*, 19(3):159–165, 2004. (cited on page 44)

Y. Liu, W. C. Hsiao, Q. Li, X. Liu, and M. Ren. Transformation of China's rural health care financing. *Social Science & Medicine*, 41(8):1085–1093, 1995. (cited on pages 45 and 46)

S. Lu, J. Su, Q. Xiang, J. Zhou, and M. Wu. Accuracy of self-reported height, weight, and waist circumference in a general adult Chinese population. *Population Health Metrics*, 14(30), 2016. (cited on page 67)

R. MacFarquhar and M. Schoenhals. *Mao's Last Revolution*. Harvard University Press, 2006. (cited on page 44)

R. Maria-Dolores and J. M. Martinez-Carrion. The relationship between height and economic development in Spain, 1850–1958. *Economics and Human Biology*, 9:30–44, 2011. (cited on page 39)

C. L. Martin and R. A. Fabes. The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, 37(3):431–446, 2001. (cited on pages 76 and 95)

M. McElroy and D. T. Yang. Carrots and sticks: Fertility effects of China's population policies. *American Economic Review*, 90(2), 2000. (cited on pages 6 and 8)

K. McGarry and X. Sun. Three generations of changing gender patterns of schooling in the people's republic of China. *ADBI Working Paper 834: Asian Development Bank Institute*, 2018. (cited on page 71)

X. Meng. *Labour market reform in China*. Cambridge University Press, 2000. (cited on pages 42 and 43)

X. Meng. Labor market outcomes and reforms in China. *Journal of Economic Perspectives*, 26(4): 75–102, 2012. (cited on page 45)

X. Meng and N. Qian. The long term consequences of famine on survivors: Evidence from a unique natural experiment using China's Great Famine. *NBER Working paper No. 14917*, 2009. (cited on page 40)

X. Meng, N. Qian, and P. Yared. The institutional causes of China's Great Famine, 1959–1961. *Review of Economic Studies*, 82:1568–1611, 2015. (cited on page 43)

C. N. Milwertz. *Accepting Population Control: Urban Chinese Women and the One-child Family Policy*. Curzon Press, 1997. (cited on pages 3 and 7)

S. L. Morgan. Richer and taller: Stature and living standards in China, 1979-1995. *The China Journal*, 44:1–39, 2000. (cited on page 40)

NBS. Comprehensive statistical data and materials on 50 years of new China, 1999. (cited on pages 42, 44, and 58)

N. Niida. Land reform and new marriage law in China. *Developing Economies*, 2(1), 1964. (cited on page 16)

M. d. Onis, C. Garza, C. G. Victora, A. W. Onyango, E. A. Frongillo, and J. Martines. The who multicentre growth reference study: Planning, study design, and methodology. *Food and Nutrition Bulletin*, 25(1), 2004. (cited on page 157)

P. P., J. D., J. R., and H. P. Distortion in self-reported height and weight data. *Journal of the American Dietetic Association*, 78(6), 1981. (cited on page 67)

E. Pahlke, J. S. Hyde, and C. M. Allison. The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, 140(4):1042–1072, 2014. (cited on pages 75 and 95)

H. Park, J. R. Behrman, and J. Choi. Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in Seoul high schools. *Demography*, 50: 447–469, 2013. (cited on pages 75, 76, 77, 79, 80, 90, 91, 92, and 94)

X. Peng. *Demographic Transition in China*. Oxford: Clarendon Press, 1991. (cited on pages 8, 10, 45, and 47)

N. Peter, P. Lundborg, S. Mikkelsen, , and D. Webbink. The effect of a sibling's gender on earnings and family formation. *Labour Economics*, 54, 2018. (cited on page 17)

I. Plante, R. d. l. Sablonniere, J. M. Nronson, and M. Theoret. Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology*, 38:225–235, 2013. (cited on pages 76 and 95)

S. H. Potter. The position of peasants in modern China's social order. *Modern China*, 9(4), 1983. (cited on page 11)

N. Qian. Quantity-quality and the one child policy: The positive effect of family size on school enrollment in China. *Gender and Development*, 2017. (cited on pages 5, 6, 16, and 17)

X. Qin, C. C. Zhuang, and R. Yang. Does the one-child policy improve children's human capital in urban China? a regression discontinuity design. *Journal of Comparative Economics*, 45:287–303, 2017. (cited on page 5)

J. A. Qureshi. Siblings, teachers, and spillovers on academic achievement. *Journal of Human Resources*, 53(1), 2018. (cited on page 25)

M. R. Rosenzweig and K. I. Wolpin. Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, 48(1), 1980. (cited on page 17)

M. R. Rosenzweig and J. Zhang. Do population control policies induce more human capital investment? twins, birth weight and China's "one-child" policy. *Review of Economic Studies*, 76, 2009. (cited on pages 5, 6, 10, 17, and 23)

M. L. Rowland. Self-reported weight and height. *The American Journal of Clinical Nutrition*, 52 (6):1125–1133, December 1990. (cited on page 67)

J. W. Salaff. Mortality decline in the people's republic of China and the united states. *Population Studies*, 27(3):551–576, 1973. (cited on page 42)

T. Scharping. *Birth control in China 1949-2000: population policy and demographic development*. Routledge, 2003. (cited on pages 4, 5, 8, 9, 10, 17, 20, 60, 122, 123, and 125)

D. Schwekendiek and J. Baten. Height development of men and women from China, South Korea, and Taiwan during the rapid economic transformation period of the 1960s-1980s. *Economics and Human Biology*, 34:169–180, 2019. (cited on pages 40 and 56)

C. Shen, J. Jin, and H.-f. Zou. Fiscal decentralization in China: History, impact, challenges and next steps. *Annals of Economics and Finance*, 13(1):1–51, 2012. (cited on page 46)

K. Silventoinen. Determinant of variation in adult body height. *Journal of Biosocial Science*, 35: 263–285, 2003. (cited on page 39)

H. Sohn. Mean and distributional impact of single-sex high schools on students' cognitive achievement, major choice, and test-taking behavior: Evidence from a random assignment policy in Seoul, Korea. *Economics of Education Review*, 52, 2016. (cited on pages 75, 76, 77, 79, 85, 90, 91, 92, and 94)

R. H. Steckel. Stature and the standard of living. *Journal of Economic Literature*, 33(4):1903–1940, 1995. (cited on pages 39 and 53)

R. H. Steckel. Heights and human welfare: Recent developments and new directions. *Explorations in Economic History*, 46(1):1–23, 2009. (cited on page 39)

M. R. Strain. Single-sex classes & student outcomes: Evidence from North Carolina. *Economics of Education Review*, 2013. (cited on page 75)

A. Sullivan, H. Joshi, and D. Leonard. Single-sex schooling and academic attainment at school and through the life course. *American Educational Research Journal*, 47(1):6–36, 2010. (cited on pages 75, 90, 92, and 94)

J. Tu. The lost generation - "Barefoot doctors" in post-reform China. *China Perspectives*, 4, 2016. (cited on page 44)

Y. Wan. Expansion of Chinese higher education since 1998: Its causes and outcomes. *Asia Pacific Education Review*, 7(1):19–31, 2006. (cited on page 71)

F. Wang. The future of a demographic overachiever: Long-term implications of the demographic transition in China. *Population and Development Review*, 37(Supplement):173–190, 2011. (cited on page 45)

X. Wang. China's macroeconomics in the 40 years of reform. In L. S. Ross Garnaut and C. Fang, editors, *China's 40 Years of Reform and Development 1978-2018*. ANU Press, 2018. (cited on pages 43 and 45)

Y. Wang, X. Wang, Y. Kong, J. H. Zhang, and Q. Zeng. The great Chinese famine leads to shorter and overweight females in chongqing Chinese population after 50 years. *Obesity*, 18 (3), March 2010. (cited on page 40)

M. K. Whyte, W. Feng, and Y. Cai. Challenging myths about China's one-child policy. *The China Journal*, 74, 2015. (cited on pages 4, 8, 9, 10, and 17)

World Bank. Gender gaps in China: Facts and figures. 2006. (cited on page 7)

Y. Xie and P. Lu. The sampling design of the China family panel studies (CFPS). *Chinese Journal of Sociology*, 1(14), 2015. (cited on page 11)

H.-C. Xin, J. Amsbary, and L. Powell. Urban health insurance reform in China. *International Conference on Economic Management and Trade Cooperation (EMTC 2014)*, 2014. (cited on page 42)

L. Xingzhu and C. Huaijie. China's cooperative medical system: Its historical transformations and the trend of development. *Journal of Public Health Policy*, 13(4):501–511, 1992. (cited on pages 42, 43, and 44)

Y. Yao and W. You. Half sky over China: Women's political participation and sex. China Center for Economic Research: Working Paper Series No. E2016009, 2016. (cited on page 16)

W.-J. J. Yeung. Higher education expansion and social stratification in China. *Chinese Sociological Review*, 45(4):54–80, 2013. (cited on page 71)

W. Yip and W. C. Hsiao. What drove the cycles of Chinese health system reforms? *Health Systems & Reform*, 1(1):52–16, 2015. (cited on pages 42 and 43)

J. Zeng, X. Pang, L. Zhang, A. Medina, and S. Rozelle. Gender inequality in education in China: A meta-regression analysis. *Contemporary Economic Policy*, 32(2):474–491, 2014. (cited on page 71)

D. Zhang and P. U. Unschuld. China's barefoot doctor: past, present, and future. *The Lancet*, pages 1865–1867, 2008. (cited on page 44)

J. Zhang. The evolution of China's one-child policy and its effects on family outcomes. *The Journal of Economic Perspectives*, 31(1):141–159, 2017. (cited on pages 4, 5, 6, 8, 10, 17, and 38)

Z. Zhang and Q. Chen. The expansion of higher education admissions and the gender equalization of higher education opportunity: an empirical study based on Chinese general social survey (CGSS2008) data. *The Journal of Chinese Sociology*, 1(1), 2014. (cited on page 71)

N. Zhu, Z. Ling, J. Lane, and S. Hu. Factors associated with the decline of the cooperative medical system and barefoot doctors in rural China. *Bulletin of the World Health Organization*, 67(4):431–441, 1989. (cited on page 43)