EMPIRICAL ESSAYS ON FAMILY AND EDUCATION ECONOMICS

A Dissertation

by

SU-CHIN HSU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Timothy J. Gronberg
Committee Members,	Li Gan
	Mark Hoekstra
	Lori L. Taylor
Head of Department,	Timothy J. Gronberg

December 2018

Major Subject: Economics

ABSTRACT

Family- and child-related topics in China have drawn wide attention from the public and scholars. A number of studies have been interested in the impact of the recent change from the one-child policy to the two-child policy in 2016 in China. The new policy raised births by 1.31 million in China in 2016 shortly. Births then decreased by 0.63 million in 2017. The government originally estimated the new policy would increase 3 million newborns annually in the five or six years following the policy change; however, the estimate did not take into account the low willingness of having a first child. The increase in 2016 fell far short of the government's target and could not compensate the decrease in the willingness of having a first child. According to fertility intention and behavior surveys conducted in China, the low willingness of having a child or having an additional child was attributed to the soaring costs of raising children. Cost estimates of raising children can provide an important insight into how children affect family economic decisions. In China, couples with two children are over-represented by a lowly-educated, rural, and sons-preferred population. This is commonly attributed to an exemption in the one-child policy: only particular couples could have a second child, and these couples were disadvantaged, lived in rural areas, and had a strong preference for sons. Using data from the Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics conducted in 2011, I employ a selection bias correction and equivalence scale method to estimate the costs of raising children that consider economies of scale enjoyed by a couple living with children. I find no difference between the costs of raising a son and a daughter for a representative one-child family. The marginal cost of raising an additional child was higher for a representative couple with a firstborn daughter and a second-born son than a couple with children of the same gender. To raise an additional child, on average, the cost was higher for an urban family than a rural one, and the cost was higher for a couple with more years of schooling than one with fewer years of education.

Besides the family planning policy, another subject-matter of concern is China's unequal access to educational resources among children from different backgrounds – rich and poor, urban and ru-

ral. Many countries have introduced conditional cash transfer programs, which provide incentives to relatively disadvantaged individuals or households when specific requirements are met. I evaluate the impact of a social experiment, the Youth Education Improvement Program, designed to improve youth's academic achievement in a neediest area through conditional cash transfers. The estimates show that the incentive had positive impacts on student performance in both standardized Chinese and mathematics examinations when using a sample with balanced characteristics. I also estimate the impacts by baseline performance and find the most significant effects of the monetary incentives for students with a relatively low baseline score while finding no significant impacts for students with a relatively high baseline score. Results using a sample with unbalanced characteristics yield no significant impacts across different baseline scores.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Professor Timothy J. Gronberg, Department of Economics, for his motivation, guidance, patience, and constant encouragement throughout my Ph.D. study.

Besides my advisor, I want to express my special appreciation to Professor Li Gan, Department of Economics, who has given so much advice and continuous support throughout the writing of my dissertation.

I thank Professor Mark Hoekstra, Department of Economics, and Professor Lori L. Taylor, Department of Public Service and Administration for taking the time to look over my work, many helpful comments, and valuable suggestions.

I would like to acknowledge the help provided by Qing He, Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics, for his kind assistance to use the data.

My sincere thanks to Professor Dennis W. Jansen, Department of Economics, for all the opportunities, help, and support he has given to me during these years.

My thanks go to my parents (especially my dearest mom), my parents-in-law, my husband, my daughter, and my son for their encouragement and never-ending support, without whose love, this dissertation would never be possible.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Timothy J. Gronberg (Chair) of the Department of Economics, Professor Li Gan of the Department of Economics, Professor Mark Hoekstra of the Department of Economics, and Professor Lori L. Taylor of the Department of Public Service and Administration.

The data analyzed for chapters One and Two were provided by the Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the Department of Economics at Texas A&M University and the E. Ralph Daniel Scholarship from Private Enterprise Research Center at Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Family- and child-related topics in China have drawn wide attention from the public and scholars. A number of studies have been interested in the impact of the recent change from the one-child policy to the two-child policy in China. Instead of studying the impact of the policy change, I investigate the costs of raising children that largely affects fertility decisions. Besides the family planning policy, another subject-matter is China's unequal access to educational resources among children from different backgrounds – rich and poor, urban and rural. My study evaluates the impact of a social experiment designed to improve youth's academic achievement in neediest areas in China through conditional cash transfers.

1.1 The Population Plan Policy

The government guidelines for population planning plays a vital role in affecting fertility decisions in China. The increasing life expectancy and declining child mortality rate accelerated China's population growth and put a strain on economic development. In the 1970s the Chinese government launched a campaign that promoted delaying marriage, a longer interval between births, and having fewer children in response to the shortage of natural resources and the problem of poverty. As a result, the total fertility rate declined from 5.5 to 3.0 births per woman before the implementation of the one-child policy. In 1978, birth planning was formally included in China's Constitution. China's one-child policy was introduced between 1978 and 1980 to further halt China's high population growth. It officially restricted the number of children per family across most provinces. Families who lived in provinces of rural western China, such as, Yunnan, Qinghai, Ningxia, and Xinjiang, were exempt from the policy restriction. Ethnic minorities and families with special circumstances were also exempted. Beginning in late 1982, local officials in rural areas were given the authority to grant approvals for having a second child if both the husband and wife were only children. This policy relaxation had a minor impact at that time due to the majority of couples being born from multi-child families. Between 1985 and 2011, the majority

of provinces successively decreed that rural families with financial or other difficulties could apply to have a second child (after a specified birth spacing) if the firstborn was a daughter. The World Bank World Development Indicators show that as of 2015 the total fertility rate in China dropped to 1.6 births per woman while the world average was 2.5 births per woman.

A further loosening of the one-child policy in 2014 allowed a family to apply for having a second a child if either the husband or wife was an only child. According to the Chinese National Health and Family Planning Commission, among the additional 11 million families that were benefited from the loosening of the policy restriction from having a second child, only about 1 million families applied to have a second child by the end of 2014. This number was half of the government's estimate of 2 million applicant families each year. In 2014, most couples who were born between 1978 and 1986 when the one-child policy was enacted (aged between 28 and 36 years old) were eligible to have a second child under the loosened policy due to them being very likely to be an only child in their families of origin. To legally have a second child, an eligible family would have applied with the government. The low number of applicants might provides some evidence that families may not have been willing to have a second child due to reasons other than the policy restriction.

The Chinese National Health and Family Planning Commission credited the one-child policy with successfully mitigating the stress of overpopulation on economic development and natural resources by preventing 400 million births as of 2014. This drop in fertility rate also brought the unique social phenomena – the imbalance in age and gender composition – in China. The unbalanced sex ratio could be explained by the implementation of the one-child policy associated with an intensified preference for a son, the diagnostic ultrasound for sex determination, and the sex-selective abortion. According to the World Bank World Development Indicators, in 2015 the Chinese sex ratio at birth was 115 boys to 100 girls compared to 107 boys to 100 girls in 1978. The ratio was much larger in some western provinces of China. These numbers show a disproportionate ratio in favor of boys when compared to the world average sex ratio at birth: 107 boys to 100 girls in 2015. The abnormal sex ratio at birth has brought about a considerable shortage of women of

marriageable age in China. According to the China National Bureau of Statistics, it is predicted there will be around 30 million more men than women of marriageable age in China by 2020.

Another problem caused by the one-child policy is the aging population. The number of workers aged 16 to 59 has begun to shrink by 3.71 million in 2014 compared to 2013 for the first time in the past two decades. The working-age population is projected to continue shrinking every year. This diminishing working-age population has to bear an enormous burden from supporting a large number of elderly. A person who was born in the generation of the one-child policy has to provide support for parents and likely to for four grandparents now, called the "4-2-1" problem. Since the total fertility rate in China was only 1.6 births per woman in 2015 according to the World Bank, the working-age population will keep shrinking. The word average was 2.5 births per woman in 2015. The replacement rate that leads to zero population growth is about two births per woman for industrialized countries and has to be greater than 2.5 births per woman for developing countries due to higher infant mortality rates. The combination of a low total fertility rate and the rapidly aging population can destabilize the Chinese economy. The direct and indirect consequences of the one-child policy have received great attention from the Chinese government.

Driven by fears that the unbalanced sex ratio and the aging population could cause social problems and hinder the economic growth, the Chinese government replaced the 35-year-old one-child policy with a two-child policy on January 1, 2016. As a result, all couples nationwide can now choose to have two children from here on forward. The objective of the new family planning policy is to balance the male-to-female ratio and the elderly-to-workforce ratio in China by stimulating the total fertility rate. Since the two-child policy was implemented, around 90 million additional families became eligible to have an additional child. The government forecasted the policy would lead to an additional 3 million births per year in the following five to six years since half of the newly eligible women were older than 40 years old and some families might continue to choose to have only one child regardless. According to China's National Health and Family Planning Commission, births increased by 1.31 million (7.9%) in 2016, which fell far short of the government's estimate. It is evident the new fertility policy has not stimulated the total fertility rate since the to-

tal fertility rate remained at 1.6 births per woman in 2017, according to the World Bank. In 2017, births decreased by 0.63 million as the increase in the second born children did not compensate the decrease in the firstborn. Additionally, the working-age population continued shrinking, falling by 5.48 million in 2017 relative to 2016 according to the figure released by the China National Bureau of Statistics. These factors led to the high sex ration level at birth in China, 115 boys to 100 girls in 2017, according to the World Bank World Development Indicators.

In addition to many factors such as, demographic, sociological, bio-medical, family income, consumption and saving, labor-force decision, and the benefit of having children, fertility decisions are primarily affected by the cost of having children. The cost of raising children is thought to be a substantial financial burden to most families, and the increased costs will decrease the number of women who give birth to a second child. The idea that China's two child-policy can induce an increase in births has been proven to be overly optimistic. The two-child policy can bring the extra number of births has been proven to be too optimistic. Although two children are widely identified as the ideal number of children for a family in China, there is evidence of a difference between this ideal number and the actual number of births that occur. This discrepancy between fertility willingness and behavior has been widely observed in the literature. Mao (2009) uses data from the Fertility Intention and Fertility Behavior Survey conducted in 2007 in Jiangsu Province in China to document the differences between fertility beliefs and result. Mao found that 2% of the women had an actual number of children greater than the number they were willing to have, 53.9% of the women displayed actual fertility behavior that matches their intentions, and 44% of the women gave birth to a fewer number of children than their willingness to have. Restricting analysis to women who already had one child and who were allowed to have the second one, the intended fertility rate was 1.47 while the actual fertility rate was only 1.01. The expected fertility rate was between 1.26 and 1.1s when future fertility plan was taken into account. Mao and Hao (2013) use data from the Fertility Intention and Fertility Behavior Survey conducted in 2010 in Jiangsu Province and suggest that of couples who were eligible to have a second child, 53.3% of women thought one child was ideal and had one child. Moreover, 6.25% identified a 2-child

family with two children as ideal and had two children while 39.91% thought a two-child family as ideal but only had one child. Zhang (2016) uses survey data conducted by the Tianjin Academy of Social Science in 2015 in the city of Beijing and documents that while 73.2% of married and unmarried respondents identified two children as ideal, only 15.4% of married respondents who already had a child considered having another one (84.6% did not plan to have a second child). Furthermore, of the 15.4% of married respondents who already had a child and were considering have another one, 3% had already become pregnant with a second one, 61.2% made plans to have a second, and 35.8% had no plans to have a second one yet. Of the 84.6% respondents who already had a child and did not consider having a second one (22% were aged more than 36 years old), 19.9% thought that one child had already been a significant financial burden or worried about the economic pressures a second one could bring Further, 17.9% turned out to not be eligible to have a second child due to the policy restriction, 15.4% identified one child as ideal, and 13.1% thought having a second child would be tiring since they were also caring for an older child. Apparently, the policy restriction is not the only key factor that affected fertility decisions for the past decade. Demographic, sociological, bio-medical, and economic characteristics, labor-force decision, and the cost and benefit of having children are all factors that may affect fertility decisions (Willis, 1973). Among all these determinants, one of the most cited reasons for having only one child is the soaring cost of raising an additional child.

Governments in many countries offer various supporting programs, subsidies, or tax credits to help families ease the financial burden of raising children. Many economic models of fertility following Becker (1960) have studied the impact of the cost of raising children on fertility decisions. There is also much literature that studies the effect of economic incentives on fertility. Buttner and Lutz (1990) observe a jump in fertility rates between 1976 and 1978 in response to the pronatalist social policy package introduced by the German Democratic Republic in 1976. Hoem (1990) and Walker (1995) examine the effect of pronatalist policies that were designed to provide financial incentives for encouraging childbearing and shorten the spacing of births in Sweden. Hyatt and Milne (1991) and Zhang et al. (1994) both use a time series model to estimate the possible impact

of the personal tax exemption for children, child tax credit, and family allowances on the total fertility rate. They find evidence that the three tax-transfer programs positively affect the total fertility pattern in Canada. Whittington et al. (1990), Whittington (1992), and Whittington (1993) find evidence that tax policies that benefit families with children did change the total fertility rate in the United States. Manski and Mayshar (2003) find that the child allowance and other related welfare programs in Israel may play an essential role in the puzzle of Israel's fertility reversion. However, the time series approach used in the literature mentioned above leaves the identification vulnerable to trends in unobserved variables. If unobserved characteristics related to childbearing decisions change over time, the time series variation may not be sufficient to identify the effect of the pronatalist policies. The following three studies use individual-level data and identify a positive effect of the pronatalist policies. Milligan (2005) finds that the introduction of a child tax subsidy in the 1990s, in the province of Quebec, Canada, had a significant, positive effect on fertility outcomes. In the analysis, Milligan compares fertility outcomes in the province of Quebec to other provinces over the same period. Cohen et al. (2013) find that financial incentives provided by government policies encouraged families to have another child by using panel data from Israel's Central Bureau of Statistic. Laroque and Salanié (2014) use data from the French Labor Force Survey and the French tax-benefit system and find that tax incentives affect fertility decision for both the first and the third child in France. In the past, a fine of 3-10 times the annual salary of a couple had to be paid to the government to have children more than permitted in China.

It seems that China's two-child policy is unable to spur births to the level estimated by the government. Estimating the costs of raising children can provide essential references to determine the method of support and the level of subsidy or tax credits in the future. The cost of raising children can be divided into two major groups: direct and indirect costs. Direct costs include additional expenditures incurred by a family due to the presence of children, such as food, clothing, child care, education, housing, etc. Indirect costs refer to opportunity costs. Examples include income forgone by parents as a result of leaving work, parents' career prospects affected, hours working declining, or changes in working types to accompany children. In my essay, I focus on

the issue of measuring the direct material cost of raising children.

In most countries, there is no policy restriction on the number of children a couple can have. However, families with two children have characteristics that are dissimilar to families with one child due to the one-child policy, which began in the 1970s. Families with two children are over-represented by lower-educated couples residing in rural areas as the one-child policy were strictly applied to couples with relatively higher earnings residing in urban areas in China. While 48.55% of the couples with one child both received more than 12 years of education, only 5.57% of the couples with two children received more than 12 years of schooling in my sample that represents the population in China. 75.07% of the families with one child lived in urban areas, and 32.74% of the families with two children resided in urban areas. The cost of two children is observed for only restricted nonrandom subgroups as a result of the one-child policy. When the probability that a particular couple has a second child under the one-child policy depends on variables that also affect the share of the second child on household's total expenditure, there may exist sample selection bias. Thus, the sample selection bias has to be considered when estimating the cost of two children. Finally, I find no difference between the costs for a representative couple for raising a boy (9.71%) or a girl (9.96%). The marginal cost of raising an additional child is higher for a representative couple with a firstborn girl and a second-born boy (33.40%) than with other gender combinations of two children (9.74% to 9.99%). I find it costs an urban couple more to raise an additional child than a rural couple and costs a higher-educated couple more than a lower-educated couple.

The essay is organized as follows. In Section 2.1, I review some of the literature on measuring the cost of raising children. Section 2.2 introduces the household behavioral models and the assumptions for estimating the cost of raising children. In section 2.3, I turn to the empirical implementation of the model. The correction of selection bias is presented in Section 2.4. Section 2.5 presents the data and empirical results. Section 2.6 concludes.

1.2 Conditional Cash Transfers

Conditional cash transfer programs were invented as social assistance tools to motivate recipients' investments in human capital with an over-arching goal of alleviating long-term poverty. The primary mechanisms of conditional cash transfer programs are: easing the short-term financial burden when changing behavior, increasing human capital accumulation, and preventing the intergenerational transmission of poverty in the long-term. In many countries, conditional cash transfer programs have been employed as policy instruments to deliver money to individuals and families on the condition that those individuals and families fulfill pre-specified requirements. These requirements are usually human capital investments, such as attending schools, performing well in examinations, or having health check-ups. Incentive payments are usually made primarily to relatively poor individuals or families for investments in human capital through education and health as with relatively poor economics resources people tend to invest fewer resources than needed. The theoretical basis for using conditional cash transfer as a motivator has been well researched. Research on the implementation and impact of conditional cash transfer programs has been well documented in many countries as well. In Latin America for example, many countries have various versions of conditional cash transfer programs.

Mexico's Progresa, launched in 1997 when Mexico was recovering from the economic crisis, was the first national conditional cash transfer program. Prospera (named initially as Progresa) was founded in 2002 with the aim of building human capital by providing incentive payments directly to low-income families. Prospera specially targeted children's and mothers' educational, health, and nutritional status. Progresa initially targeted only poor rural areas but later expanded its coverage to poor urban areas. The program went from covering 300,000 families in 1997 when the pilot program was launched to including about five million families in 2002 when the program was revised. The program works by providing direct cash transfers every other month to female heads of families as long as specified requirements are met. The conditional cash transfers vary by a child's gender and grade level to provide a stronger incentive for households to send girls to school. Many studies have evaluated the impact of Prospera in Mexico, see Schultz (2000a),

Schultz (2000b), Skoufias (2001), Schultz (2004), and de Brauw and Hoddinott (2011). Among these studies, Schultz (2004) shows that the program had induced an increase in enrollment rates in 1999 for relatively poor third- to ninth-graders and an increase in the number of schooling by 0.66 years, 0.72 years of additional education for girls and 0.64 years of education for boys.

In Brazil, conditional cash transfer programs have been the most common social assistance tool since 1988. Several pre-existing conditional cash transfer programs were combined to form Bolsa Família in 2001. The program initially benefited 6.7 million families and is currently the largest conditional cash transfer program in the world, covering more than 25% of the Brazilian population. Bolsa Família works by providing incentive payments made to poor households for sending children to school and having children vaccinated. The objectives of Bolsa Família include eliminating child labor, raising school attendance rates, improving mothers' and children's health status, and poverty. Conditional cash transfers are offered to female heads of the relatively low-income families given the compliance of the pre-specified conditions. Compared to Mexico's Prospera, Bolsa Família is applied to citizens more generally. Glewwe and Kassouf (2012) evaluate the effects of Bolsa Família on enrollment, dropout rates, and grade promotion. They find that the program raised enrollment by around 5.5% for first- to fourth-graders and by 6.5% for fifth- to eighth-graders. In addition, dropout rates declined by 0.5 percentage points for first- to fourth-graders graders and by 0.4 percentage points for fifth- to eighth-graders; grade promotion rates increased by 0.9 percentage points in grades 1-4 and 0.3 percentage points in grades 5-8. de Brauw et al. (2015) examines the impact of Bolsa Família and observes significantly positive effects on enrollment and grade progression for only girls using survey conducted in 2005 and 2009. Girls' enrollment rate was increased by 8.2 percentage points with their grade progression rate increasing by 22.5 percentage points for girls aged 15-17 years old and by 14.6 percentage points for girls aged 6-14 years old in rural areas. In urban areas, the program raised girls' progression by 18.0 percentage points for those aged 15-17 years old but did not affect those aged 6-14 years old. Both Brazil's Bolsa Família and Mexico's Prospera programs attempts to alleviate the short-term financial burden, increase human capital accumulation, and thus prevent the long-term

inter-generational cycle of poverty.

Following the implementation of Mexico's Prospera and Brazil's Bolsa Família, most of the countries in Latin America borrowed elements from Brazil and Mexico and adopted their own conditional cash transfer programs as policy instruments to deliver social assistance. In other parts of the world, many countries have also embraced conditional cash programs. Bangladesh, Cambodia, Indonesia, Malawi, Morocco, Pakistan, South Africa, Turkey, the United States, among others, have also adopted such monetary incentives as the policy instrument to alter recipients' behavior. In Bangladesh and Cambodia, conditional cash transfers have been employed to alleviate gender inequality in education. Chile and Turkey specifically target gender inequality in education for poor households. Conditional cash transfers in Sub-Saharan Africa are used to improve youth's health and social protection. The Opportunity NYC-Family Rewards, a privately funded conditional cash transfer program, is the first comprehensive conditional cash transfer program in New York city that aims at improving children's education, family health care, and parents' work outcomes for low-income families by offering cash incentives. Conditional cash transfers have become popular and have been thought of as a useful tool to accomplish the goals of encouraging investment in human capital and alleviating poverty.

A prerequisite for the policymaker when designing a conditional cash transfer program is to determine the recipients of the incentive payments in the first place. If all people are qualified to the conditional cash transfer program, people who do not need the incentive payments may also receive the benefits, referred to as the inclusion error. However, if there is a pre-established criterion for determining the qualification of receiving conditional cash transfers, people more in need of the incentive payments may be excluded from the program, resulting in the exclusion error. Thus, a well-designed conditional cash program should target beneficiaries who are the neediest and deserving to receive the incentive payments in the population. The most widely adopted selection method is based on geographic targeting and means testing. Conditional cash transfer programs usually first target the population in areas with relatively low income compared to other geographical regions and also account for the socioeconomic status.

Conditional cash transfer programs may offer incentives to students (parents), teachers, or both students and teachers. Muralidharan and Sundararaman (2011) examine the impacts of the teacher incentive program that rewards teacher for improvement on student performance in students' test scores in rural India. Schools were randomly assigned as either one of the treatment groups or the control group. They find that test scores in math and language tests increased by 0.27 and 0.17 standard deviations respectively for students in the schools where teachers were offered incentives. Glewwe et al. (2003) also evaluate the effects of a randomized experiment that offered fourth- through eighth-grade teachers in primary schools for student performance on districtwide exams. Students in the treatment schools on average obtained 0.144 standard deviations higher test scores in the second year than those in the control schools. Angrist and Lavy (2009), Kremer et al. (2009), Fryer (2011), Barrow and Rouse (2013) study the effects of effects of monetary incentive paid to students and find evidence that incentives positively affect students performance and study behaviors. Behrman et al. (2015) examine the effects of three performance-based incentives designs: incentives provided to students only, to teachers only, and to both students and teachers. They find the most significant effects for groups in which students and teachers were both offered incentives.

My study is based on the Youth Education Improvement Program (YEIP), a conditional cash transfer program that provided students (and teachers) monetary incentives in Mabian Yi Autonomous County, located in the Sichuan Province in China. In Mabian Yi Autonomous County, 47.51% of the residents are the Yi people, and many of the households do not take youth's education seriously due to various economic and geographic reasons. The rationale for utilizing conditional cash transfers is that offering monetary incentives tied to multiple measures of performance encourages parents and students, especially students who performed poorly, to invest in education to improve academic achievement and encourage teachers to help with students. The ultimate goal of the program is to increase human capital accumulation and thus prevent the inter-generational transmission of poverty.

I estimate the impact of YEIP and show that the incentive demonstrated a positive effect on stu-

dent performance in both standardized Chinese and mathematics examinations when the treatment group did not differ from the control group in individual, household, and class characteristics. In the sample, students are classified into four groups according to their performance in the previous baseline semester. Performance in the baseline semester is taken into account in estimation to better measure the impact of the conditional cash transfer program on student performance. My estimates further show the most significant average impact on student performance in both Chinese and mathematics for students with a relatively low baseline grade when the control group and the treatment do not differ from each other in characteristics.

The essay is organized as follows. Section 3.1 reviews some literature on conditional cash transfers. In section 3.2, I introduce the YEIP (the sample design and implementation, the survey, the examinations, and the incentives). The empirical results are presented in section 3.3. Section 3.4 concludes.

2. THE COST OF RAISING CHILDREN

2.1 Literature Review

Among the related literature, Kornrich and Furstenberg (2013) examine how investments in children vary across different periods using detailed expenditure data on clothing, child care and education expenditures, and toys, hobbies, and playground equipment from the Consumer Expenditure Survey data (CES) by the United States Bureau Statistics. The main result thay find is that between the early 1970s and late 2000s, families shifted from heavier investments in male children to heavier investments in female children. Families additionally shifted from investing more in children's teen years to investing more when children are under age 6 and when in their mid-20. These results illustrate changes in gender and age differences in the cost of raising children. Lino et al. (2017) calculate expenditures on children in 2015 for the United States Department of Agriculture (USDA) by using the CES. Excluding clothing, child care and education expenses, the CES does not provide details on who incurred expenses and the purposes for the expenses. Allocation rules computed from other sources are required to calculate the estimated expenses to child-bearing. The 2015 USDA food plans based on food intake recommendations and cost of food for various categories and quantities are used to compute the share of total household food expenses for children. The share of household out-of-pocket health care expenses spent on children obtained from the United States Department of Health and Human Services 2012 Medical Expenditure Panel Survey is applied to calculate the health care expenses for children. The United States Department of Transportation study estimates that 75% of household transportation expenses is family-related, and the family-related expenses are allocated in equal proportions among household members. Miscellaneous expenses are also allocated equally among family members. Housing expenses on children are estimated from the average cost of an additional bedroom by first regressing housing expenses on the number of bedrooms in a house. After the assignment is calculated, the cost of raising children is estimated by regressing the expenditure on children on total

income, family size, and the age of the younger children. Household demographic characteristics are not considered in the calculation.

The current survey data available usually does not specify who receives what; thus, it is difficult to determine individual goods and services when computing expenditures on children directly. In the absence of individual-level data and in the attempt to consider household economic behavior between families of different compositions, an indirect method that utilizes the survey data to compare the total expenditures of childless families and those of families with children who have a comparable standard of living is introduced to estimate the cost of raising children. The method that computes the extra expenditure required to maintain the standard of living of a family unaffected by the presence of children also faces drawbacks. One challenge in estimation comes from making households of different compositions comparable. Additionally, an increase in family size as a consequence of the presence of children generates economies of scales in consumption and thus changes the shadow prices and the shadow budgets for all family members.

Equivalence scale that measures the ratio of the costs of achieving the same level of the material living standard under two different family compositions is commonly utilized in the literature to estimate the cost of having children. The cost of raising children can be obtained by first, computing equivalence scale by taking the ratio of expenditures of households with and without children that achieve the same living standard. The supplementary portion of total expenditure is considered as the cost of having children. Engel (1895) is the earliest study that derives an equivalence scale, using the share of total expenditure spent on food as the indicator of living standard. This implies that two households enjoy the same level of the living standard if they consume an equal share of total expenditure on food. This share declines as total household expenditure increases and usually rises with household sizes. A decline in share is regarded as an increase in the standard of living, and vice versa. Let s_0 be the family size of the base family and x_0 the total household expenditure, another family of larger size s_1 has a total expenditure of x_1. The two families achieve the same level of the living standard if they both consume (q/x_1) of total expenditure on food. Thus, the equivalence scale defined as the ratio of x_1 and x_0, i.e. x_1/x_0, gives the scalar by which total ex-

penditure of the larger family s_1 must be multiplied to achieve the same living standard as the base family s_0. This approach implicitly assumes that all goods consumed in a household are impacted by the presence of the new child in the same way; that is, increments in all goods rise in the same proportion as food expenditure does.

Rothbarth (1943) suggests using the expenditure level spent on goods exclusively consumed by adults, such as adult clothing, tobacco, and alcohol, instead of the share of food consumption as the measure of the living standard for different family compositions. The Rothbarth approach focuses on children's impact on the amount spent on goods exclusively consumed by adults, i.e., adult-specific goods. In contrast, the Engel approach focuses on effects on the total expenditure share of jointly consumed goods of food. Under the Rothbarth approach, adults are assumed to act in a selfish manner and thus wish to keep their consumption of adult-specific goods unchanged to maintain the same living standard. If two families spend an equal amount on goods exclusively consumed by adults, they attain the same living standard in terms of the level of adult consumption. Assuming the base family, s_0, has a total household expenditure of x_0, and the larger family s_1 has a total expenditure of x_1. The two families achieve the same living standard if they consume the same amount of q on adult-specific goods. Thus, the ratio of total expenditures x_1 and x_0, i.e., x_1/x_0, with equal expenditure spent on adult-specific goods represents the equivalence scale that enables the larger family, s_1, achieves the same standard of living as the base family, s_0, does.

Adult-specific goods are required to be separable from children's consumption for estimating the cost of raising children. Deaton et al. (1989) use the 1980-1981 household survey data in Spain, the Spanish Encuestas de Presupuestos Familiares, to test the assumption that adult-specific consumption is demographically separable from children for different categories of adult-specific goods candidates and age group combinations. Demographic separability in this paper implies that the presence of children aged 0-13 years old should induce no direct demand for adult-specific goods. Overall, most of the adult-specific goods candidates in the paper are accepted as demographically separable from the presence of children. Gronau (1991) tests the validity of the separability assumption that the presence of children has the same impact on all adult goods. Us-

ing the United States 1960 and 1972 CES data on adult-specific goods, including men's clothing, women's clothing, tobacco, and alcoholic beverages, all successfully pass the test, individually and combined. Surveys of the main approaches employed in the literature, including Engel, Rothbarth, Prais and Houthakker, Barten, and Gorman, are shown in Browning (1992) and Nelson (1993). Among all approaches, the Rothbarth approach is one of the most employed. Gronau (1988) also verifies that if parents' material living standard is used as a benchmark, the Rothbarth approach is the only one available method that complies with the assumption and can be used to derive an equivalence scale. Further, the existence of adult-specific goods are required to identify them.

The original Rothbarth approach does not account for the possibility of scale economies if family size increases with the presence of children. Bargain and Donni (2012) follow the approach used by Lewbel and Pendakur (2008) that incorporates economies of scale into the estimation of parents' and child's shares of total expenditure with an extended collective model using the 2000 French Household Budget Survey data. Differing from traditional literature on collective models given by Lewbel and Pendakur (2008) and Browning et al. (2013) (where children are treated as consuming together with their mother), in Bargain and Donni (2012), both parents and children have preferences represented by indirect utility functions. Bargain and Donni include single men, single women, childless couples, and couples with one child to identify the effect of scale economies on each individual. They also identify the individual difference between a couple in their responses to the presence of a first child. Men's clothing and women's clothing are used as adult-specific goods for husband and wife respectively. They find that the expenditure made for children is 23% of total expenditure of a family for a girl and 26.5% for a boy. However, the estimated cost of children computed by their approach – the percentage of the household expenditure of a childless couple must be compensated in order to maintain the level of utility of the parents unaffected after the presence of the first child – is very small at around 5.3% of total expenditure for a boy and about 0.4% for a girl due to the economies of scale. Dunbar et al. (2013) is another paper that employs a collective model in which children also have their own utility. They use the 2004-2005 second Malawi Integrated Household Survey data for their analysis. The esti-

mation sample includes observations of couples with one to four children. Economies of scale in expenditure is not estimated in this paper though. They find that families make a large share of household expenditure on children, ranging from 13.5% to 29.3% of total expenditures for families with one to four children. Although the share of total expenditure increases with the number of children, the average share of total expenditure each child obtains decreases when the number of children increases. Results show boys are favored over girls; a child's share of total expenditure falls by from 4.8% to 6.5% if all the children are girls. In this essay, detailed information on the expenditure spent for adult clothing and other consumer products in conjunction with information of household wealth from the 2011 Chinese Household Finance Survey make the identification of resource shares, economies of scales, and equivalence scales possible.

2.2 The Model

In this section, the household behavioral models are specified. I follow the approach and identification assumptions proposed by Lewbel and Pendakur (2008) for childless couples and later Bargain and Donni (2012) for couples with one child to estimate the equivalence scale, the cost of raising one or two children, and the marginal cost of raising an additional child. There are three types of families: childless couples, couples with one child, and couples with two children.

2.2.1 Household Resource Shares between Members and Members' Budget Share Equations

The following sections describe models for a childless couple and a couple with one child. The case of a couple with two children can be analyzed similarly. A couple and a child in a household are denoted as $m = a$ and k respectively.

2.2.1.1 A Childless Family

A childless family with a vector of individual characteristics z_a consumes a vector of private or public goods $c_a = (c_a^1, \ldots, c_a^Q)$ and has a monotonically increasing, strictly quasi-concave, and continuously twice differentiable direct utility function expressed by $U_a(c_a, z_a)$ that measures the material standard of living. Given the log total expenditure constraint x and a vector of log prices

p, the childless family solves the following utility maximization problem to determine the level of consumption over a vector of private or public goods

$$\max_{c_a} \; U_a(c_a; z_a)$$
$$\text{subject to } \; p'c_a = x.$$

The solution to this utility maximization problem is denoted by

$$c_a = c_a^*(p, x, z_a).$$

The indirect utility function that is three times differentiable and independent of the household demographic structure corresponding to U_a can be defined by

$$V_a(p, x, z_a) = U\left(c_a^*(p, x, z_a); z_a\right).$$

The budget share of a couple for good c_a^q at a price of p^q can be obtained using Roy's identity

$$\omega_a^q(p, x, z_a) = -\frac{\partial V_a(p, x, z_a)/\partial \log p^q}{\partial V_a(p, x, z_a)/\partial \log x},$$

for $q = 1, \ldots, Q$.

2.2.1.2 A Family with One Child

In a family consisting of a couple and a child, the family members $m = a$ and k face the log total expenditure constraint x and a vector of log prices p. The couple receives a share of total expenditure represented by η_a, according to their bargaining power summarized by the differentiable sharing function. The child receives a share of total expenditure denoted by η_k. Furthermore, the increase in the number of household members as a result of the presence of a child creates economies of scale. The cost savings enjoyed by the couple from economies of scale in consumption because of the presence of a child is described by s_a.

To achieve the identification goal, assumptions are summarized as in Lewbel and Pendakur (2008) and Bargain and Donni (2012).

Assumption 1. The sharing functions, $\eta_m(p, z_a, z_k)$ for $m = a$ and k, are differentiable, independent of the total household expenditure, comprised between zero and one, and summed up to one for the sake of simplicity, i.e. $\eta_a(p, z_a, z_k) = 1$ in families consisting of a childless couple and $\eta_k(p, z_a, z_k) = 1 - \eta_a(p, z_a, z_k)$ in families consisting of a couple and a child.

Assumption 2. For a couple with a child, there exists a scalar-valued, differentiable function, $s_a(p, z)$, such that the indifference curve of the couple satisfy the following condition:

$$U_a = V_a(p, x + \log \eta_a(p, z_a, z_k) - \log s_a(p, z_a), z_a),$$

for any level of log expenditure allocated to the couple, $x + \log \eta_a(p, z_a, z_k)$. A child is assumed to always live with parents in a household, $s_k = 1$.

The log total expenditure is allocated between the couple and the child in a family, and then utility functions are maximized given a Lindahl (1919) type shadow price vector and the shadow budget. The decision process can be described by the following weighted sum of the utility functions.

$$\begin{aligned}
\max_{c_a, c_k, c} \quad & \eta_a(p, z_a, z_k) U_a(c_a; z_a) + (1 - \eta_a(p, z_a, z_k))\, U_k(c_k; z_k) \\
\text{subject to} \quad & \log(p' c_a) = x + \log \eta_a(p, z_a, z_k) - \log s_a(p, z_a), \\
& \log(p' c_k) = x + \log(1 - \eta_a(p, z_a, z_k)), \\
& c = c_a s_a(p, z_a) + c_k, \\
& p' c = x,
\end{aligned}$$

The solution to this weighted sum utility maximization problem is

$$
\begin{aligned}
c_a(p, x, z_a, z_k) &= c_a^* \left(p, x + \log \eta_a(p, z_a, z_k) - \log s_a(p, z_a), z_a\right), \\
c_k(p, x, z_a, z_k) &= c_k^* \left(p, x + \log(1 - \eta_a(p, z_a, z_k), z_k\right), \\
c(p, x, z_a, z_k) &= c_a \left(p, x, z_a, z_k\right) s_a(p, z_a) + c_k(p, x, z_a, z_k).
\end{aligned}
$$

In a family consisting of a couple and a child, given resource shares, the fraction of resource share spent on good c_m^q can be derived by applying Roy's identity

$$
\omega_m^q(p, x, z_a, z_k) = -\frac{\partial V_m(p, x_m - \log s_m(p, z_m), z_m)/\partial p^q}{\partial V_m(p, x_m - \log s_m(p, z_m), z_m)/\partial x_m}\bigg|_{x_m = x + \log \eta_m(p, z_a, z_k)},
$$

for $m = a, k$ and $q = 1, \ldots, Q$. Developing the derivatives can lead to

$$
\begin{aligned}
\omega_m^q(p, x, z_a, z_k) &= \lambda_m(p, z_m) + \omega_m^q(p, x_m - \log s_m(p, z_m), z_m) \\
&= \lambda_m(p, z_m) + \omega_m^q(p, x - \log I(p, z_a, z_k), z_m),
\end{aligned}
$$

where $\lambda_m(p, z_m) = \partial \ln s_m(p, z_m)/\partial p_q$ is the price elasticity of $s_m(p, z)$ with respect to the q^{th} product, which is equal to zero for a child since $s_m(p, z_m) = 1$ for a child living in a household. The equivalence scale by which the total expenditure must be multiplied so that a couple with a child can reach the same living standard as being a childless couple is

$$
\mu_e(p, z_a, z_k) = \frac{s_a(p, z_m)}{\eta_a(p, z_a, z_k)}.
$$

It becomes explicit that $s_a(p, m_{,}) = \eta_a(p, z_a, z_k)$ when all the consumption is purely public and $s_a(p, z_m) = 1$ when all the goods are private.

The material cost of raising one child can be computed by

$$
C_1(p, z_a, z_k) = \mu_e(p, z_a, z_k) - 1.
$$

The material cost of raising two children, $C_2(p, z)$, can be derived by a similar procedure.

2.3 Empirical Implementation

2.3.1 Identification Strategy

The goal in this intermediate stage is to identify the sharing and the scale economies functions from the observations of childless couples, couples with one child, or couples with two children. As long as the observations are observed when the vector of log prices contains very little or no variation, the log price factor p can be taken out from equations. Given each member's budget share and the allocated level of log total household expenditure, the household budget share for good q can be computed by

$$W^q(x, z_a, z_k) = \sum_m \eta_m(z_a, z_k)\omega_m^q(x, z_a, z_k),$$

for $q = 1, \ldots, Q$.

The critical identifying idea is that, after conditioning on observed demographic variables, the difference in expenditure patterns between a childless couple and a couple with one child are attributed entirely to joint consumption (economies of scale in consumption) and resource allocation. Identification only requires the existence of an adult-specific good for the couple in a household.

For a childless couple, the household budget share spent on good q is equal to the budget share of the couple

$$W_0^q(x, z_a) = \omega_a^q(x, z_a),$$

for $q = 1, \ldots, Q$.

For a couple with one child, the household budget share spent on adult-specific good q_a is

$$W_1^{q_a}(x, z_a, z_k) = \eta_a(z_a, z_k)\left[\lambda_a^{q_a}(z_a) + \omega_a^{q_a}(x - \log I_a(z_a, z_k), z_a)\right],$$

as the adult-specific good is consumed by only the couple. The function $\lambda_a^{q_a}(z_a)$ can be discarded

by having first-order derivative with respect to the log total expenditure x:

$$\nabla_x W_1^{q_a}(x, z_a, z_k) = \eta_a(z_a, z_k)\nabla_x \omega_a^{q_a}(x - \log I_a(z_a, z_k), z_a),$$

Differentiating the above equation again. The second-order derivative with respect to total expenditure x is

$$\nabla_{xx} W_1^{q_a}(x, z_a, z_k) = \eta_a(z_a, z_k)\nabla_{xx} \omega_a^{q_a}(x - \log I_a(z_a, z_k), z_a).$$

Taking the ratio of the above two equations:

$$\frac{\nabla_x W_1^{q_a}(x, z_a, z_k)}{\nabla_{xx} W_1^{q_a}(x, z_a, z_k)} = \frac{\nabla_x \omega_a^{q_a}(x - \log I_a(z_a, z_k), z_a)}{\nabla_{xx} \omega_a^{q_a}(x - \log I_a(z_a, z_k), z_a)} = \triangle_a^{q_a}(x - \log I_a(z_a, z_k), z_a).$$

The left-hand side and $\triangle_a^{q_a}(\cdot, z_a)$ are known from the budget share of childless couples. Thus, the equivalence scale, $\eta_a(z_a, z_k)/s_a(z_a)$, can be uniquely identified provided that $\triangle_a^{q_a}(\cdot)$ is not periodic in its first argument. Then, sharing function, $\eta_a(z_a, z_k)$, can be identified from the first-order derivative equation, and the scaling function, $s_a(z_a)$, can be obtained from the definition of $\eta_a(z_a, z_k)/s_a(z_a)$. Lastly, the identification of $\lambda_a(z_a)$ follows from the household budget share spent on the adult-specific good q_a.

The sharing function and scaling function for couples with two children can be identified by utilizing observations of childless couples and couples with two children.

2.3.2 Functional Form

The households considered in this study are childless couples and couples with one child or two children. The model for estimation can be written as the following system of equations. The budget share equation of member $m = a$ and k for good q is

$$\omega_m^q(x_m, z_m) = \bar{\gamma}_m^q + z_m \gamma_m^q + \delta_m^q (x_m - z_m \tau_m) + \zeta_m^q (x_m - z_m \tau_m)^2,$$

for $m = a$ and k, $q = 1, \ldots, Q$, and $\bar{\gamma}_m^q, \gamma_m^q, \delta_m^q, \tau_m$, and ζ_m^q are parameters. For a couple, the characteristics entering $z_m\gamma_m^q$ for adults include husband's or wife's age and education, whichever is greater, dummies for urban resident and total household income greater than household total expenditure, and those entering $z_m\tau_m$ include husband's or wife's age and education, whichever is greater. For children, the characteristics entering both $z_m\gamma_m^q$ and $z_m\tau_m$ include age and gender of the first child for families with one child and age of the first child and a dummy for a gender combination of a female firstborn and male second-born for families with two children.

For a childless couple, the household budget share is just the budget share of the couple spent on good q specified above plus an error term

$$\tilde{W}_0^q(x, z) = \omega_a^q(x, z_a) + \varepsilon_0^q.$$

For a family with one child, the household budget share spent on good q is

$$\tilde{W}_1^q(x, z) = \eta_a(z)[\lambda_a^q(z) + \omega_a^q(x - \log I_a(z), z_a)] + (1 - \eta_a(z))\omega_k^q(x + \log(1 - \eta_a(z)), z_k) + \varepsilon_1^q.$$

The budget share equations for member $m = a$ and k are as specified. The sharing function is a logistic form:

$$\eta_m(z) = \frac{\exp(\bar{\beta}_m + z_m\beta_m)}{\exp(\bar{\beta}_a + z_a\beta_a) + \exp(\bar{\beta}_k + z_k\beta_k)},$$

where $\bar{\beta}_m$ and β_m for $m = a$ and k are parameters. The economies scaling function is

$$\log s_a(z) = \bar{\alpha}_a + z_a\alpha,$$

where $\bar{\alpha}_a$ and α_a are parameters. Lastly, the price elasticity of $s_a(z)$ is restricted to be constant

$$\lambda_a^q(z) = \bar{\lambda}_a^q.$$

For a family with two children, the household budget share spent on good q is similar to the

share of a fmaily with one child

$$\tilde{W}_2^q(x,z) = \eta_a(z)[\lambda_a^q(z) + \omega_a^q(x + \log I_a(z), z_a)] + (1 - \eta_a(z))\omega_k^q(x + \log(1 - \eta_a(z)), z_k) + \varepsilon_2^q,$$

where $\eta_a(z)$ represents the budget share for a couple with two children.

Households in the data are indexed by n, and the numbers of childless families, families with one child, and families with two children in the data are denoted by N_0, N_1, and N_2, respectively. Let $\mathbf{W}_{c,n}$ be the $(K-1)$ vector of observed budget shares for the first $(K-1)$ goods consumed by household n with the number of children denoted by c, and $\hat{\mathbf{W}}_{c,n}(\theta)$ be the corresponding $(K-1)$ vector of predicted budget shares for some parameter vector θ. The vector of residuals is given by $\varepsilon_{c,n}(\theta) = \mathbf{W}_{c,n} - \hat{\mathbf{W}}_{c,n}(\theta)$. If $\hat{\varepsilon}_{c,n} = \varepsilon_{c,n}(\hat{\theta_0})$, where $\hat{\theta_0}$ is any initial consistent estimation for the vector of parameters, the estimated covariance matrix can be defined by

$$\hat{\mathbf{V}}_c = \frac{(\hat{\varepsilon}_{c,n})(\hat{\varepsilon}_{c,n})'}{\mathbf{N}_c}.$$

The model will be estimated by the iterated SURE method with the criterion

$$\min_{\theta} \sum_{c=0}^{1} \sum_{n=1}^{N_c} (\varepsilon_{c,n}(\theta))'(\hat{\mathbf{V}}_c)^{-1}, (\varepsilon_{c,n}(\theta))$$

which gives a new value $\hat{\theta_1}$ for the estimates. I then iterate with new estimates until the covariance matrix to obtain the final estimates.

2.4 The Correction of Selection Bias

The number of children is usually the result of a family's decision according to a family's preference. The expectation of the marginal cost of raising a second child can be computed by taking the difference between the expected costs of raising two children and one child when the fertility decision is determined without intervention outside the family. However, policy restriction in China played an important role in determining the number of children in a family until the one-child policy ended in 2016. Between 1978 and 2015, only couples who met the requirements set by

the government could apply to have a second child without penalty. Couples who were not allowed to have a second child could pay the penalty according to their own economic and social economic status to have an additional child. Qualifying to have an additional child does not necessarily imply a family would desire another child at the same time, there existed non-qualified families who would be willing to pay the penalty to have a second child. The general stopping rule related to son preference is that a family will continue to have an additional child until the ideal number of male children is reached. In the case of China, a family is more likely to have a second child when the first-born is a daughter because of the strong son preference. This is particularly true for families with one daughter facing financial or other difficulties in rural areas. As a result, among the two-child families, a relatively high proportion of those families have a firstborn daughter and second born son. The gender imbalance is more significant in rural areas where social and family economic status are lower. Moreover, in my sample, a large proportion of couples with two children received less than 12 years of education (91.63%); and a large proportion lived in rural China (60.00%). A potential selection bias issue exists when the cost of two children is estimated since I only observe the cost of raising two children for a very restricted nonrandom group of families who are more likely to have strong son preference, receive fewer years of schooling, and live in poor rural areas.

The probability that couples in the sample of the two-child families is a function of 1) years of education, 2) areas of residence due to the exceptions to the one-child policy for some particular groups, and 3) preference for a son. The household total expenditure for a family strongly depends on a couple's years of schooling and geographic location. Estimating causal effects will be a problem if selection on observables (education and residence area) exists. In particular, certain values of total household expenditure are over-represented due to their relationship with education and areas of residence permit. Horvitz and Thompson (1952) is a very early paper that presents a, now frequently used, inverse probability weighting scheme that accounts for different proportions of observations within a strata of a population. When the sampling probability is known (derived from the sampling population of a population), the inverse of this probability is used to weight the

observations. I adopt the inverse probability weighting method to correct the sample selection bias that results from over-representation of certain types of families. While 51.60% of the couples with one child obtained more than 12 years of schooling and lived in urban areas, only 7.44% of the couples with two children achieved more than 12 years of schooling and lived in urban areas. 28.72% and 32.56% of the one-child and two-child couples respectively received less than 12 years of schooling and lived in urban areas. 3.35% and 0.93% of one-child and two-child couples respectively received more than 12 years of education and lived in rural areas. Finally, while only 16.33% of the couples with one child received less than 12 years of schooling and lived in rural areas, 59.07% of the couples with two children received less than 12 years of schooling and lived in rural areas. Thus, if I reweight the impact that lower-educated rural couples and well-educated urban couples have in the estimation of the cost of raising two children, I would alter the distribution of types in the two-child family sample to match the distribution of types in the one-child family sample. Inverse probability weighting method weights each family's impact by the inverse of the probability that the family appears in the two-child family sample.

Before I can correct the selection bias that accrues due to the exception of some particular groups of the one-child policy, the probability of appearing in the two-child family sample has to be estimated first. This is accomplished by fitting a Probit regression of appearing in the two-child family sample given the standard normal distribution assumption on the error terms. Let k_2 be the indicator of whether a couple has a second child, $1\{\bullet\}$ be an indicator function, and ϵ_{k_2} follows a normal distribution. With s represents the segment for the combination of dummies for years of education greater than 12 years and urban residence permit. The probability that a couple have a second child can be obtained by fitting the following Probit regression

$$k_2 = 1\{s + \epsilon_{k_{2o}}\}.$$

The inverse probability needed to match the distribution of the two-child family sample to the distribution of the one-child family sample is the inverse of the fitted probability obtained above.

To understand the problem of sample selection bias on unobservables, for example gender preference, a set of two equations of interest are involved. The first equation is the choice between having a second child or not. Let k_2 be indicators of whether a couple has the second child or not. With a set of exogenous variables z and an indicator function $1\{\bullet\}$, a family chooses to have the second child if the net value of having the second child, $z\beta - \varepsilon_{k_{2u}}$ is greater than 0. The second equation expresses the household budget share spent on good q for a family with two children. The two equations are written as follows

$$k_2 = 1\{z\beta > \epsilon_{k_{2u}}\}$$

$$\tilde{W}^q_{2,n} = \begin{cases} \eta_a(z)[\lambda^q_a(z) + \omega^q_a(x - \log I_a(z), z_a)] \\ \qquad +(1-\eta_a(z))\omega^q_k(x + \log(1-\eta_a(z)), z_k) + \varepsilon^q_2 & \text{if } k_2 = 1 \\ - & \text{otherwise.} \end{cases}$$

(Heckman, 1979) sets the correlation between the independent variable and the error term as an omitted variable in the ordinary least square moment condition. That is,

$$\begin{aligned} E(\tilde{W}^q_{2,n}|x, z, k_2 = 1) =& \eta_a(z)[\lambda^q_a(z) + \omega^q_a(x - \log I_a(z), z_a)] \\ & + (1-\eta_a(z))\omega^q_k(x + \log(1-\eta_a(z)), z_k) + E(\varepsilon^q_2|x, z, k_2 = 1). \end{aligned}$$

Furthermore, using the conditional density of the standard normal distributed ε^q_2,

$$E(\varepsilon^q_2|x, z, k_2 = 1) = \frac{-\phi(z\beta)}{\Phi(z\beta)},$$

which is called the inverse Mills ratio, where ϕ and Φ are the probability and cumulative density functions of the standard normal distribution. As a result, the moment condition that holds in the

sample is

$$E(\tilde{W}^q_{2,n}|x,z,k_2=1) = \eta_a(z)[\lambda^q_a(z) + \omega^q_a(x - \log I_a(z), z_a)] + (1-\eta_a(z))\omega^q_k(x + \log(1-\eta_a(z)), z_k) + \delta\frac{-\phi(z\beta)}{\Phi(z\beta)},$$

where $\delta = \sigma_{12} \neq 0$ as error terms are correlated.

2.5 Data and Empirical Results

2.5.1 Data and Sample

The data comes from the China Household Finance Survey (CHFS), conducted in 2011. The survey was designed by the Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics and includes 8,438 households. Households were asked to recall their expenditures in each category in the previous month or previous year. In the estimation of budget share equations, $Q = 10$ commodities are included: adult clothing, food at home, food out, household operation, leisure, transportation, housing, education, other categories, and children's clothing (the omitted good in the system). The selection criterion for households included in the sample is described below. Only nucleus families without a child and with one child or two children aged less than 18 years old in which both husband and wife are working adults aged 20-59 are included. The final sample consists of 307 childless households, 686 one-child households, and 215 households with two children, that is, 1,208 households in total.

Table 2.1 and Table 2.2 summarize the final sample. The descriptive statistics tables provide an overview of the issues that have to be addressed. Due to the exemption in the one-child policy that allowed rural households with financial difficulties to have a second child if their firstborn was a girl and the prevailing strong son preference, 64% of the firstborns of the two-child households are daughters. This provides clear evidence that families with a firstborn daughter were more likely to be allowed and were more willing to have another child. Families that valued boys over girls were more likely to practice diagnostic ultrasound for sex determination and abortion of girls if their ideal number of male children had not reached despite doctors being forbidden from exposing

the gender of unborn babies. As further evidence of the strong preference for sons, 61% of the second-born children in the two-child families are male. Among the two-child families, 42% of the families has a firstborn daughter and a second-born son, 42% has two children of the same gender, and only 16% has a firstborn son and a second-born daughter. Couples with two children on average received fewer years of schooling and earned a lower income. In addition, a large proportion of two-child families live in the middle (40%), western (28%), or rural region of China where sex ratios are relatively higher and circumstances are different from those in east China. Family economic behavior is affected by family size, parental education, and living location. Thus, estimating the cost of two children by using a subset of families that were allowed to have two children can produce distorted results.

The average share of household expenditure spent on adult clothing decreases as the household size increases due to the presence of children. This result is especially presented when a family goes from having one child to having two children. Families with one child and childless families spent the smallest and the largest fraction of total expenditure respectively on food at home. Families with two children and those with one child spent the smallest and the largest proportion on dining out respectively. The share of household expenditure spent on household operation decreased as the family size increased. Families with two children were those spent the smallest share of household expenditure on the leisure-related category. Families with one child spent the largest share of household expenditure on leisure. One-child families on average spent the largest proportion of expenditure on transportation but spent the smallest proportion on housing. Families with two children on average invested a relatively larger proportion than families with one child in the education-related category, albeit the difference was small. The average share of house expenditure spent on children clothing increases as the number of children increases, the increment was not big though.

2.5.2 Empirical Results

The coefficients related to the sharing and scaling equations are reported in Table 2.3. The first result is that the older a couple is, the greater the magnitude of scale economies enjoyed by

Table 2.1: Descriptive Statistics of the Sample

	All		Childless		One Child		Two Children	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Husband age	40.25	8.18	46.07	11.45	38.21	5.62	38.42	4.87
Wife age	38.53	8.20	44.60	11.63	36.31	5.41	36.92	4.55
Husband eudcation	11.17	3.99	10.28	4.33	12.35	3.68	8.68	2.74
Wife eudcation	10.28	4.63	8.64	5.51	11.83	3.96	7.67	2.98
Urban	0.66	0.47	0.51	0.50	0.80	0.40	0.40	0.49
Provincial level city	0.15	0.36	0.17	0.38	0.18	0.39	0.03	0.17
East	0.50	0.50	0.53	0.50	0.53	0.50	0.33	0.47
Middle	0.27	0.44	0.25	0.43	0.23	0.42	0.40	0.49
West	0.24	0.43	0.22	0.42	0.23	0.42	0.28	0.45
Houseowner	0.92	0.27	0.91	0.29	0.92	0.27	0.94	0.24
Carowner	0.20	0.40	0.10	0.31	0.27	0.44	0.12	0.33
Government employee	0.29	0.45	0.22	0.41	0.40	0.49	0.05	0.22
Log income	10.51	1.09	10.20	1.21	10.80	0.96	10.03	1.00
Log expenditure	10.43	0.95	9.90	1.10	10.73	0.78	10.26	0.85
Income $\geq$ expenditure	0.66	0.47	0.68	0.47	0.69	0.46	0.53	0.50
1^{st} kid	0.75	0.44	.	.	1.00	0.00	1.00	0.00
2^{nd} kid	0.18	0.38	.	.	.	.	1.00	0.00
1^{st} kid gender (1=boy)	0.54	0.50	.	.	0.59	0.49	0.36	0.48
2^{nd} kid gender (1=boy)	0.61	0.49	.	.	.	.	0.61	0.49
1^{st} kid age	10.58	4.91	.	.	9.82	5.03	13.01	3.56
2^{nd} kid age	7.05	3.81	.	.	.	.	7.05	3.81
1^{st} daughter & 2^{nd} son	0.42	0.50	.	.	.	.	0.42	0.50
Births spacing	5.96	3.29	.	.	.	.	5.96	3.29
Area kids sex ratio	118.61	6.51	119.03	6.82	117.53	6.04	121.46	6.61
Sample size	1,028		307		686		215	

Table 2.2: Budget Shares

	All		Childless		One Child		Two Children	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Adult clothing	0.057	0.056	0.064	0.073	0.059	0.051	0.037	0.037
Food home	0.288	0.184	0.319	0.216	0.271	0.164	0.299	0.188
Food out	0.064	0.088	0.053	0.090	0.075	0.086	0.048	0.087
Household operation	0.151	0.100	0.173	0.176	0.143	0.079	0.141	0.101
Leisure	0.042	0.062	0.045	0.044	0.048	0.063	0.022	0.032
Transportation	0.083	0.142	0.057	0.096	0.095	0.157	0.079	0.142
Housing	0.083	0.171	0.110	0.203	0.065	0.143	0.106	0.197
Education	0.066	0.098	0.012	0.057	0.080	0.103	0.094	0.101
Others	0.148	0.155	0.164	0.174	0.142	0.143	0.146	0.161
Children clothing	0.015	0.018	.	.	0.019	0.017	0.023	0.021
Sample size	1,208		307		686		215	

a couple. This is especially true for a couple living with two children. Regarding the distribution of resources between a couple and their child/children, the couple's maximum level of education has a positive effect on their own share of household resource in a one-child family, but a negative impact on a two-child family. Taking into account economies scale, the cost of raising a child declines with a couple's maximum level of education while the cost of raising two children rises with a couple's maximum level of education. Living in urban decreases the share of household resources and the magnitude of scale economies that a couple receives in a two-child household. The children's share of household resources is negatively correlated with the age of the first child in a two-child family. Once the first child attends school, the share of total expenditure devoted to the first child decreases due to the nearly free, compulsory 9-year education in China. In contrast to Bargain and Donni (2012) and Dunbar et al. (2013), gender differences are not shown in the one-child family. Nonetheless, a child's share of household resources significantly increases when a two-child family has a female firstborn and a male second-born.

The estimated shares of household resources and cost of raising children for the representative couple (aged 38 or younger with no more than 12 years of schooling and living in urban areas)

Table 2.3: Parameter Estimates of Sharing and Scaling Equations

	One Child		Two Children	
	Estimate	S.D.	Estimate	S.D.
Translations of total expenditure (s_a):				
Constant	-0.0036	0.0037	-0.0130	0.0018
Max age	-0.0068	0.0052	-0.0160	0.0100
Max education	0.0048	0.0090	0.0150	0.0184
Group (Income geq Expenditure)	-0.0250	0.0134	0.0165	0.0193
Urban	0.0460	0.0324	0.0246	0.0158
Shares of total expenditure (η_n):				
Variables entering couple's exponential function:				
Max age	0.0028	0.0071	0.0065	0.0125
Max education	0.0403	0.0099	-0.0953	0.0498
Group (Income geq Expenditure)	0.0356	0.0382	-0.0045	0.0232
Urban	-0.0754	0.0634	-0.0241	0.0184
Variables entering children's exponential function:				
Constant	-0.0072	0.0026	0.0200	0.0036
1^{st} kid age	0.0101	0.0100	-0.1345	0.0409
1^{st} kid gender (1=boy) / 1^{st} daughter & 2^{nd} son	-0.0169	0.0123	0.1875	0.0625

Note. –
Standard deviations are in parentheses.

are shown in Table 2.4 and Figure 2.1. The shares of total expenditure devoted to a child are 32.65% for a boy and 32.81% for a girl in the one-child households; the costs of raising one child are 9.71% for a boy and 9.96% for a girl. A couple with a female firstborn and a male second-born allocates a larger share of household resources to children than a couple with other gender combinations of two children (especially one with two children of the same gender), and they are 50.01% and 42.02% respectively. The costs of raising children are 43.36% for a couple with a female firstborn and a male second-born and 19.70% for one with other gender combinations of two children (especially one with two children of the same sex). Furthermore, the marginal cost of raising a second-born is 9.99% if a family already has a firstborn boy. This is very close to the marginal cost of having a second-born girl if the couple already has a firstborn daughter. The

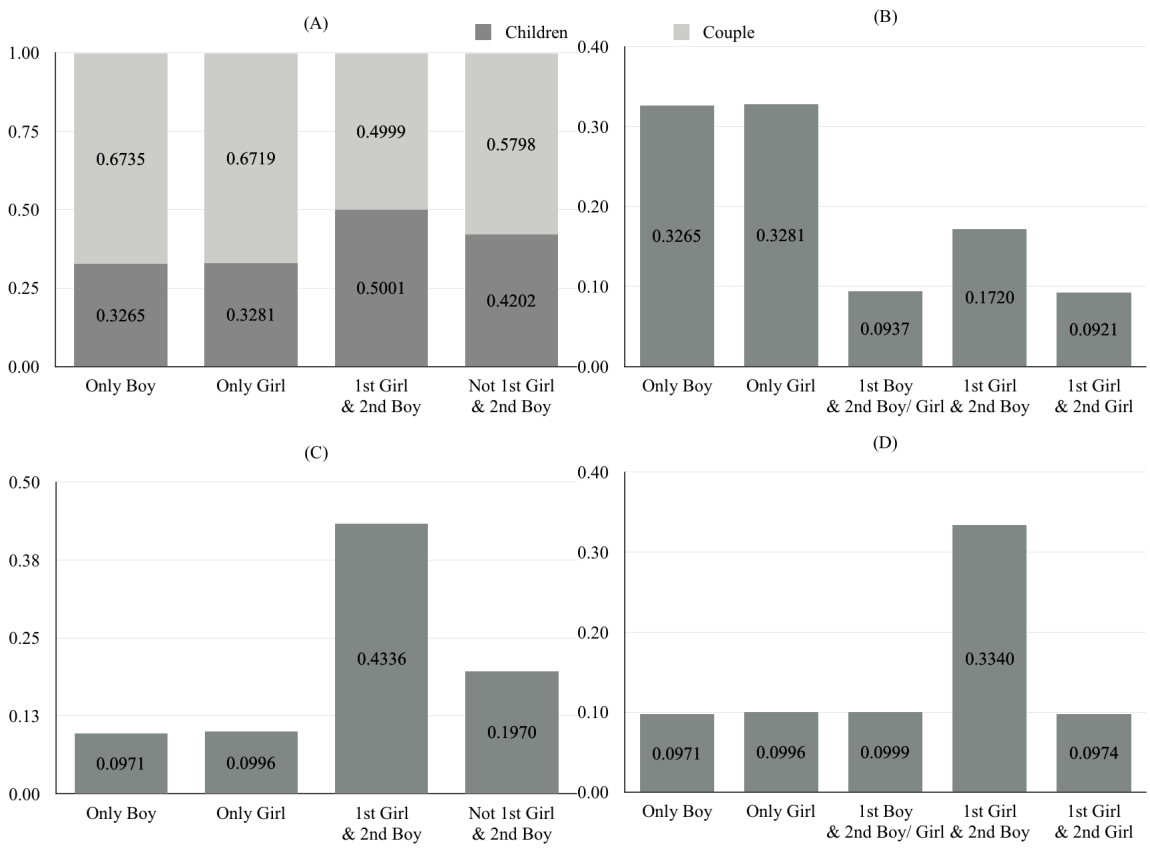

Figure 2.1: (A) Distribution of Household Resources; (B) Marginal Share of Household Resources Devoted to an Additional Child; (C) Cost of Raising Children; (D) Marginal Cost of Raising an Additional Child for the urban representative couple

interesting result here is that the marginal cost of raising an additional child jumps to 33.40% when the couple already has a girl followed by having a boy. The marginal cost of raising an additional child is very similar to the cost of raising the first child except for the case in which a couple has already had a firstborn girl and subsequently has a second-born boy. The difference in the marginal costs of raising an additional child provides clear evidence for the conjecture that sons are preferred to daughters.

Turning to the rural area, the results in Table 2.5 and Figure 2.2 show that the representative couple (aged 38 or younger with no more than 12 years of schooling) spend 33.70% and 33.57% of household resources on a male child and a female child respectively. The costs of raising a child are 8.99% for a male child and 8.59% for a female child. The couple with a firstborn girl and a

Table 2.4: Estimated Shares of Household Resources and Cost of Raising Children Aged 12 or Younger for the Urban Representative Households

	One Child		Two Children	
	Expected Value	S.D.	Expected Value	S.D.
Couple's share (boy/ 1^{st} girl & 2^{nd} boy)	0.6735	0.0141	0.4999	0.1014
Couple's economies scale (boy/ 1^{st} girl & 2^{nd} boy)	0.7383	0.0169	0.6822	0.0523
Children's share (boy/ 1^{st} girl & 2^{nd} boy)	0.3265	0.0141	0.5001	0.1014
Cost of children (boy/ 1^{st} girl & 2^{nd} boy)	0.0971	0.0445	0.4336	0.3889
Couple's share (girl/ not 1^{st} girl & 2^{nd} boy)	0.6719	0.0154	0.5798	0.1128
Couple's economies scale (girl/ not 1^{st} girl & 2^{nd} boy)	0.7382	0.0152	0.6610	0.0315
Children's share (girl/ not 1^{st} girl & 2^{nd} boy)	0.3281	0.0154	0.4202	0.1128
Cost of children (girl/ not 1^{st} girl & 2^{nd} boy)	0.0996	0.0436	0.1970	0.3050

Note. –
The representative couple are aged 38 or younger with no more than 12 years of schooling living in urban area.
Standard deviations are computed by drawing 300 bootstrap replications.

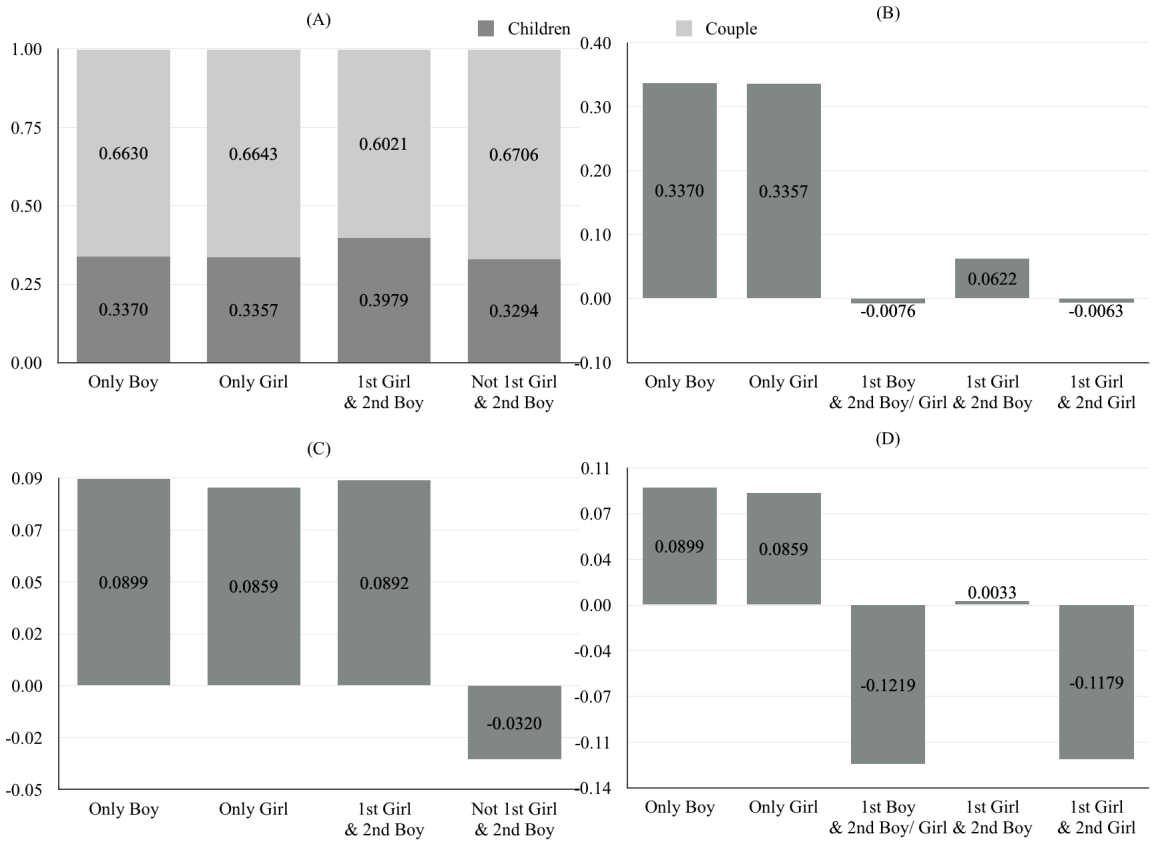

Figure 2.2: (A) Distribution of Household Resources; (B) Marginal Share of Household Resources Devoted to an Additional Child; (C) Cost of Raising Children; (D) Marginal Cost of Raising an Additional Child for the rural representative couple

second-born allocates a large share of resources to their children than the couple with two children of the same sex, and they are 39.79% and 32.94%. The costs of raising children are 8.92% for a couple with a female firstborn and a male second-born and -3.20% for raising two children with the same gender.

I further compare the differences in the total costs and marginal costs between urban and rural representative couple in Figure 2.3. The results in this essay show the marginal cost of raising an additional child is much lower for the couple living in a rural area than the urban counterpart.

As mentioned above, a couple's maximum level of education is an important factor that affects the distribution of household resources and alters the cost of raising children. Table 2.6, Table 2.7, Figure 2.4, and Figure 2.5 provide estimates of the shares of household resources and the costs

Table 2.5: Estimated Shares of Household Resources and Cost of Raising Children Aged 12 or Younger for the Rural Representative Households

	One Child		Two Children	
	Expected Value	S.D.	Expected Value	S.D.
Couple's share (boy/ 1^{st} girl & 2^{nd} boy)	0.6630	0.0131	0.6021	0.0719
Couple's economies scale (boy/ 1^{st} girl & 2^{nd} boy)	0.7222	0.0153	0.6453	0.0363
Children's share (boy/ 1^{st} girl & 2^{nd} boy)	0.3370	0.0131	0.3979	0.0719
Cost of children (boy/ 1^{st} girl & 2^{nd} boy)	0.0899	0.0401	0.0892	0.1638
Couple's share (girl/ not 1^{st} girl & 2^{nd} boy)	0.6643	0.0133	0.6706	0.0750
Couple's economies scale (girl/ not 1^{st} girl & 2^{nd} boy)	0.7209	0.0217	0.6362	0.0246
Children's share (girl/ not 1^{st} girl & 2^{nd} boy)	0.3357	0.0133	0.3294	0.0750
Cost of children (girl/ not 1^{st} girl & 2^{nd} boy)	0.0859	0.0503	-0.0320	0.1851

Note. –
The representative couple are aged 38 or younger with no more than 12 years of schooling living in rural area.
Standard deviations are computed by drawing 300 bootstrap replications.

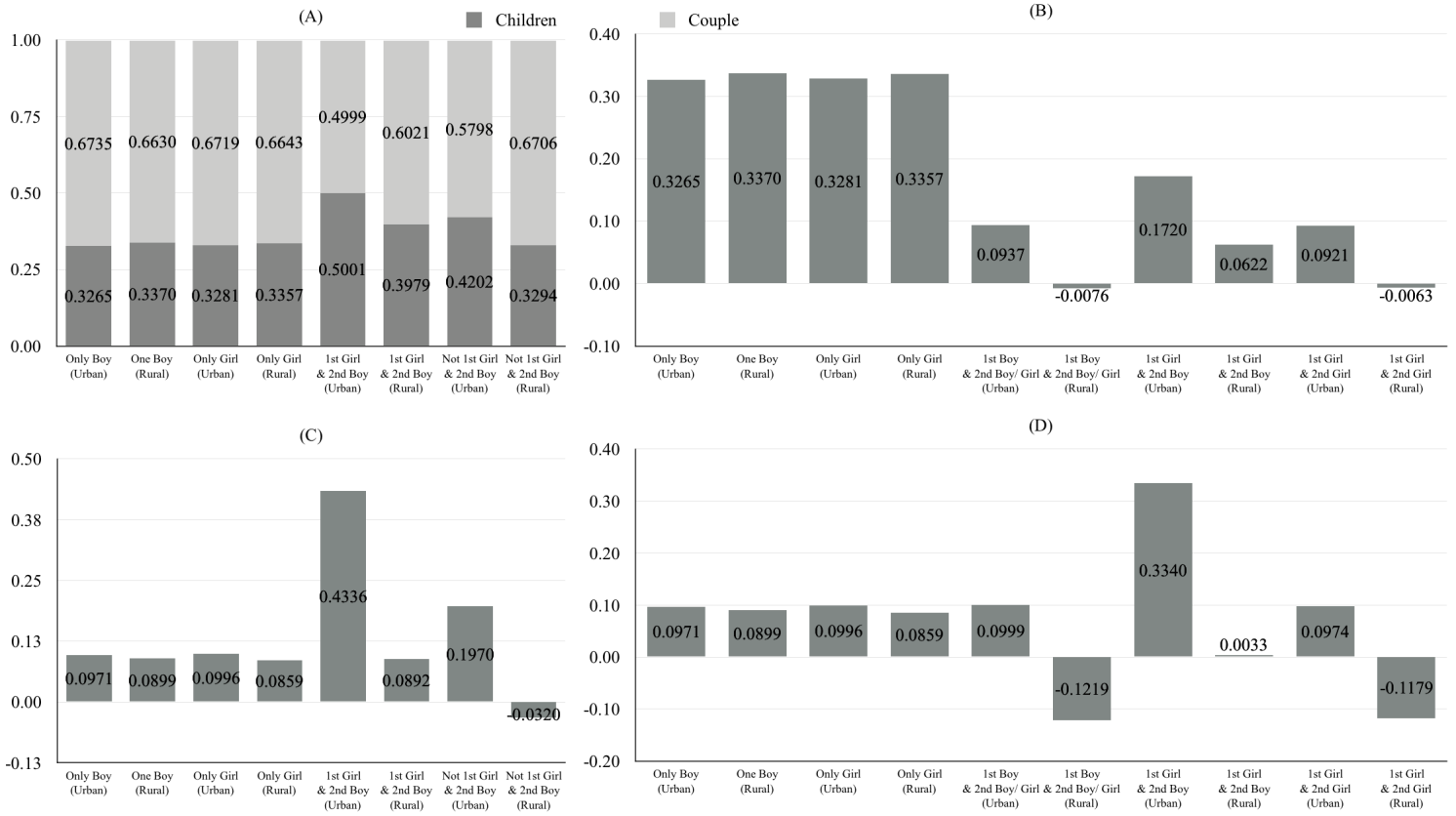

Figure 2.3: (A) Distribution of Household Resources; (B) Marginal Share of Household Resources Devoted to an Additional Child; (C) Cost of Raising Children; (D) Marginal Cost of Raising an Additional Child for the representative couple

of raising children for the representative couple. The results show a couple receives more than 9 years of schooling, spend smaller shares of household resources on the an only child (29.40% for a boy and 29.38% for a girl) than a couple that receives less 9 years of schooling (34.01% for a boy and 34.72% for a girl). The cost of raising one child is lower for the couple that receives more than 9 years of schooling (2.25% for a boy and 2.26% for a girl) than for the couple receives less than 9 years of schooling (12.73% for a boy and 14.45% for a girl). For the couple with two children, higher education level implies larger shares of household resources distributed to the children. Couples with more than nine years of schooling spend on average 53.56% on a composition of a female firstborn and a male second-born versus 43.57% for two children with the same gender. The costs of raising two children for these composition are 49.64% and 24.32% respectively. Couples with less than nine years of schooling on average spend 41.74% on a female firstborn and a male second-born and 35.30% on two children with the same sex; the costs of raising two children are 23.45% and -0.34% respectively. The shares of household resources and costs of raising children for a lower-educated couple versus a well-educated couple are compared and contrasted in Figure 2.6. The results show it costs the lower-educated couple more than the higher-educated couple to raise a child. In addition, the well-educated couple tends to have a higher marginal cost of raising an additional child than the lower-educated counterpart.

2.6 Conclusion

Departing from Bargain and Donni (2012) and Dunbar et al. (2013), the results in this essay suggest that families spent similar shares of total expenditure on a male and female child if couples had only one child. Moreover, the costs of raising a boy and a girl after considering economies of scale were not different from each other. Despite these results, the marginal cost of raising an additional child was much higher for families with a firstborn daughter and a second-born son than those with two children of the same gender. Additionally, to raise an additional child the marginal cost was higher for urban couples than rural couples, and it was also higher for couples with more years of schooling than for those with less education.

Concerns about population decline in many countries has encouraged some governments to

Table 2.6: Estimated Shares of Household Resources and Cost of Raising Children Aged 12 or Younger for the Urban Lower-Educated Representative Households

	One Child		Two Children	
	Expected Value	S.D.	Expected Value	S.D.
Couple's share (boy/ 1^{st} girl & 2^{nd} boy)	0.6599	0.0086	0.5826	0.1236
Couple's economies scale (boy/ 1^{st} girl & 2^{nd} boy)	0.7436	0.0182	0.6718	0.0556
Children's share (boy/ 1^{st} girl & 2^{nd} boy)	0.3401	0.0086	0.4174	0.1236
Cost of children (boy/ 1^{st} girl & 2^{nd} boy)	0.1273	0.0414	0.2345	0.4451
Couple's share (girl/ not 1^{st} girl & 2^{nd} boy)	0.6528	0.0098	0.6470	0.0703
Couple's economies scale (girl/ not 1^{st} girl & 2^{nd} boy)	0.7468	0.0152	0.6385	0.0092
Children's share (girl/ not 1^{st} girl & 2^{nd} boy)	0.3472	0.0098	0.3530	0.0703
Cost of children (girl/ not 1^{st} girl & 2^{nd} boy)	0.1445	0.0351	-0.0034	0.1084

Note. –
The representative couple are aged 38 or younger with less than 9 years years of schooling living in urban area.
Standard deviations are computed by drawing 300 bootstrap replications.

Table 2.7: Estimated Shares of Household Resources and Cost of Raising Children Aged 12 or Younger for the Urban Higher-Educated Representative Households

	One Child		Two Children	
	Expected Value	S.D.	Expected Value	S.D.
Couple's share (boy/ 1^{st} girl & 2^{nd} boy)	0.7060	0.0193	0.4644	0.0330
Couple's economies scale (boy/ 1^{st} girl & 2^{nd} boy)	0.7210	0.0187	0.6876	0.0371
Children's share (boy/ 1^{st} girl & 2^{nd} boy)	0.2940	0.0193	0.5356	0.0330
Cost of children (boy/ 1^{st} girl & 2^{nd} boy)	0.0225	0.0519	0.4964	0.2340
Couple's share (girl/ not 1^{st} girl & 2^{nd} son)	0.7062	0.0175	0.5643	0.1159
Couple's economies scale (girl/ not 1^{st} girl & 2^{nd} son)	0.7214	0.0166	0.6662	0.0327
Children's share (girl/ not 1^{st} girl & 2^{nd} son)	0.2938	0.0175	0.4357	0.1159
Cost of children (girl/ not 1^{st} girl & 2^{nd} boy)	0.0226	0.0457	0.2432	0.3181

Note. –
The representative couple are aged 38 or younger with more than 9 years of schooling living in urban area.
Standard deviations are computed by drawing 300 bootstrap replications.

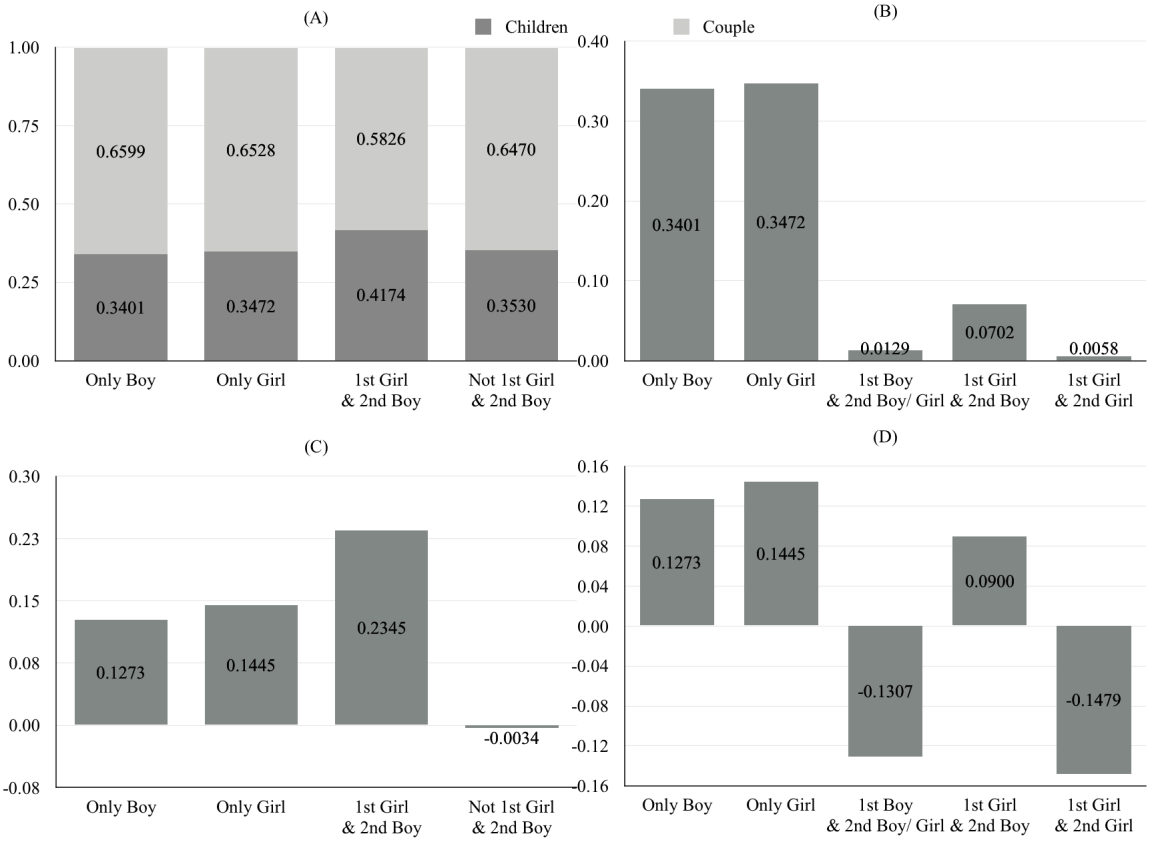

Figure 2.4: (A) Distribution of Household Resources; (B) Marginal Share of Household Resources Devoted to an Additional Child; (C) Cost of Raising Children; (D) Marginal Cost of Raising an Additional Child for the urban lower-educated representative couple

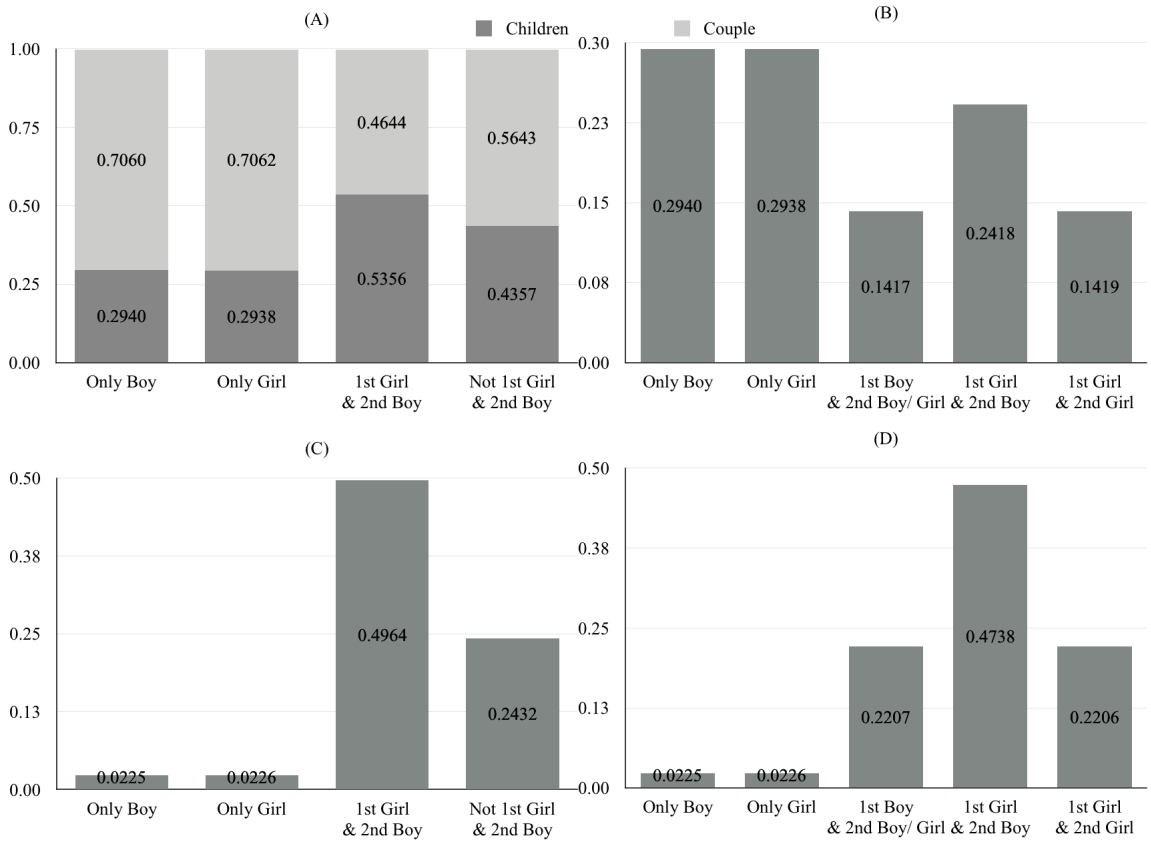

Figure 2.5: (A) Distribution of Household Resources; (B) Marginal Share of Household Resources Devoted to an Additional Child; (C) Cost of Raising Children; (D) Marginal Cost of Raising an Additional Child for the urban higher-educated representative couple

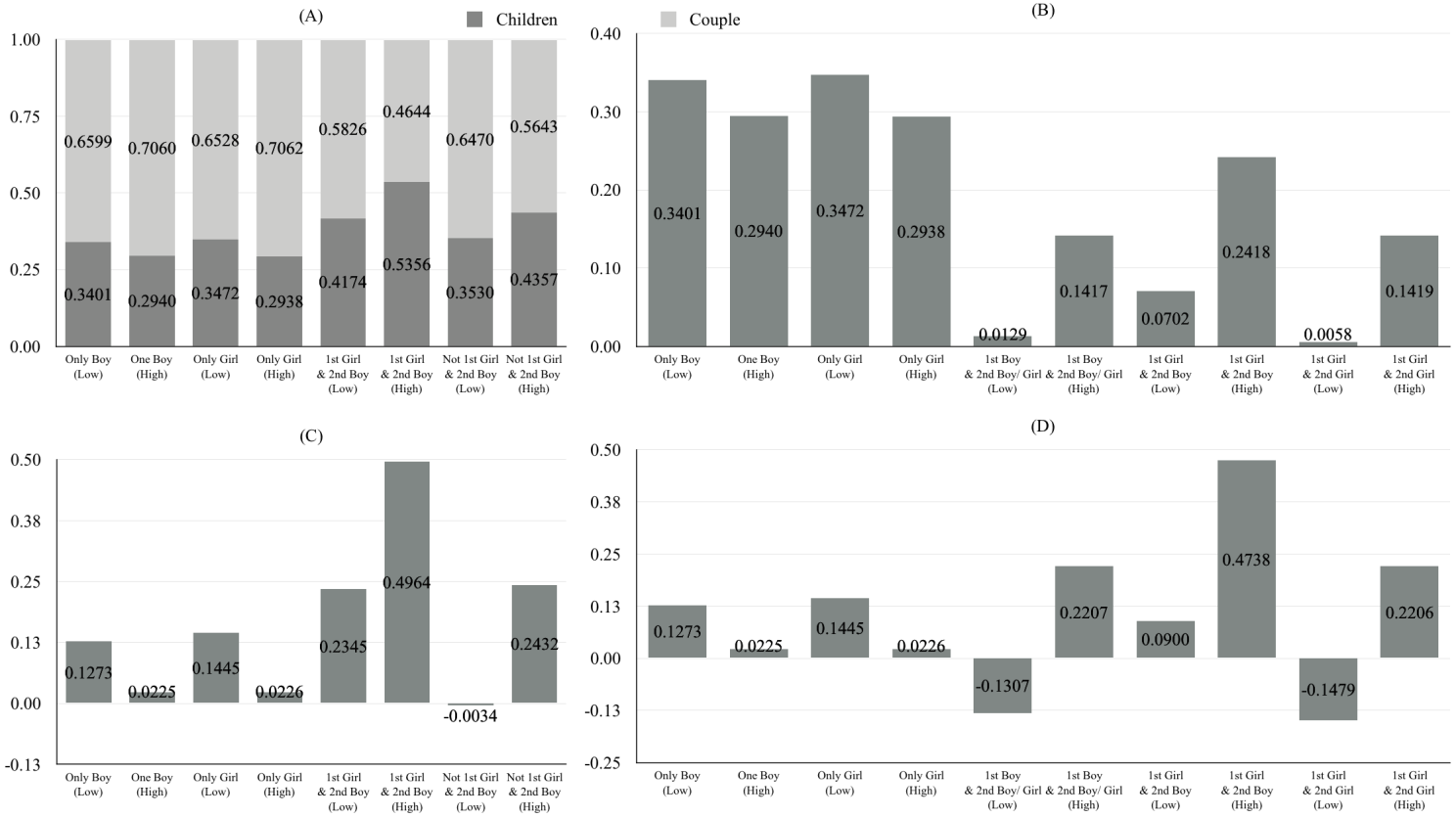

Figure 2.6: (A) Distribution of Household Resources; (B) Marginal Share of Household Resources Devoted to an Additional Child; (C) Cost of Raising Children; (D) Marginal Cost of Raising an Additional Child for the urban representative couple

adopt pronatalist policies in order to provide households with children special benefits. Some of these countries include Canada, France, Germany, Israel, Sweden, and the United States. Many of these policies are designed to increase fertility by providing financial incentives that reduce the cost of raising children. Whittington et al. (1990) find elasticities of fertility with respect to the cost of raising a child other than the eldest one ranges from 0.91 to 4.12, using the personal exemption records from 1913 to 1984 in the United States. Milligan (2005) obtains a strong, positive, and robust effect of the Allowance for Newborn Children policy that was implemented between 1988 and 1997 on fertility in Canada. The estimated elasticity of fertility with respect to the cost of raising an additional child was estimated to be 4.09. This result translates to imply a 1% decrease in cost of raising an additional child increased the total fertility rate by 4.09 births per 1,000 women. Cohen et al. (2013) find that a 1% increase in the child subsidy leads to an increase of 4.96 in total fertility rate, using Israel's child subsidy records from 1999 to 2005. The literature suggests that policies that change the cost of raising an additional child may be an effective government policy instrument to influence the total fertility rate. My estimates provide a useful reference for designing the appropriate policy instruments to alter fertility decisions.

3. CONDITIONAL CASH TRANSFERS AND STUDENT ACHIEVEMENTS: RESULTS FROM A SOCIAL EXPERIMENT IN CHINESE PRIMARY SCHOOLS

3.1 Literature Review

In this section, I review some of the literature on conditional cash transfers. Angrist and Lavy (2009) examine the impact of a monetary incentive program in Israel that offered high school students in low-achieving schools cash incentives for passing a series of national examinations in the core and elective subjects in every grade between tenth and twelfth. Forty nonvocational high schools with a certain range of the 1999 Bagrut passing rates were selected and randomly assigned to be either treatment group or the control group. The program increased Bagrut passing rates particularly for girls with a predicted marginal passing rate. The increase in the passing rates translated into an increase in postsecondary enrollment. Kremer et al. (2009) study the effects of a merit-based scholarship program for sixth-grade girls in Kenya. The program provided a grant for school fees (paid to school) and a grant for school supplies (paid to families) for two academic years. The monetary incentives were offered to girls who scored well on academic examinations. Schools were randomly assigned to treatment and control groups. The authors find evidence that the incentives increased test scores for girls and surprisingly raised scores for non-targeted boys. In addition, the largest improvements were observed in the second baseline test quartile, a group thought to have relatively low probabilities of winning an award.

Fryer (2011) evaluates the effects of various incentive programs on student achievement in three cities, Chicago, Dallas, and New York City. In Dallas, second-grade students were encouraged to read books and rewarded for passing a short quiz on them. In Chicago, students are paid for grades received in five courses. In New York City, fourth- and seventh-grade students were rewarded for performance in 10 standardized tests. Fryer finds no impact of financial incentives on student achievement in Chicago and New York City, but possibly positive effects of investment on English-speaking student achievement in Dallas. He concludes that incentives tied to inputs might

be more appropriate for students in these schools. Barrow and Rouse (2013) study the effects of performance-based scholarship programs on student behavior. The programs were offered to different target groups, students in their last year of high school and postsecondary students, with lengths varying from one to four semesters. They find that students who qualified for incentives devoted more time to educational activities and less time to non-educational activities such as work and leisure.

Behrman et al. (2015) examine the effects of three performance-based incentives designs: incentives provided to students only, to teachers only, and to both students and teachers. Eighty-eight Mexican high schools were randomly assigned to three different treatment groups and a control group. They observe the most substantial treatment effects for groups in which both students and teachers were offered incentives. Levitt et al. (2016) test the impact of a randomized field experiment that provided various monthly performance-based incentives. Ninth-graders from two schools in an academically low-performing school district or their parents received monetary rewards for student educational achievement. A 2×2 design was used with the first type of treatment being provided to either the student or parent and the other type being a fixed rate or a lottery, yielding four treatment groups. In the student treatment group, students received the monetary incentives while in the parent group, parents received the incentives. In the fixed rate treatment group, students who met the monthly achievement standard for attendance, behavior, grades, and test scores qualified for a $50 reward. In the lottery treatment group, students who met the monthly achievement standard qualified for a lottery in which there was a 10% probability of winning an award of $500. While they find a modest overall effect, a large and significant effect was observed for students on the threshold of meeting the achievement standard.

There is a debate as to whether extrinsic incentives induce an adverse effect on intrinsic motivation. Proponents of using incentives in behavioral interventions believe that monetary incentives are effective in encouraging people to invest more effort to achieve the desired outcome; however, opponents argue that, in the longer-term, extrinsic incentives may crowd out intrinsic motivations that originally result in desired consequences. The argument is that salient extrinsic incentives

may produce the desired consequence in the short-term, and intrinsic motivations can be weakened as people may pursue the desired outcome less eagerly and thus lower effort on pursuing the achievement (Gneezy et al., 2011). Bettinger (2012) evaluates the impact of a pay-for-performance program on standardized tests, including mathematics, reading, social science, and science, administered to 3^{rd} to 6^{th} grade students in Coshocton, Ohio. Bettinger shows that the program had a positive effect on students' mathematics scores but did not significantly affect the scores of students in other subjects. He additionally finds that the impact on mathematics score was most significant for students who already performed well previously. Moreover, when examining the test scores a year after the incentives were provided, he does not observe a difference in the scores between the control and treatment group. Hence, the removal of the extrinsic incentives reduced the intrinsic motivation.

Visaria et al. (2016) study the effects of a monetary reward for attending a required number of school days on attendance rates and test scores in mathematics and science for third-grade students. They find that the incentive for attending school increased attendance rates but did not affect scores significantly. Two or three months after the incentive was discontinued, students with high baseline attendance had similar attendance and test scores regardless if they were previously in the control or treatment groups. For students with low baseline attendance, attendance and test scores were decreased 2-3 months after the removal of the incentive. List et al. (2018) evaluate the effect of monetary incentives on test performance for students in grades K-8 in the short- and long-run. For the high-stakes non-incentivized test, students and tutors worked together in preparation. A week later, a separate, incentivized test that measured similar knowledge and skills as the high-stakes test was administered. The incentives were provided to either students, parents, tutors, students and parents, or all three of them. They find that the financial incentives demonstrated a positive impact on the incentivized test no matter who received the incentives. Effects for the non-incentivized test were not observed immediately but were found after one year.

3.2 The YEIP

Mabian Yi Autonomous County, under the administration of Leshan city, is a relatively poor county located in southwest China's Sichuan Province. The population of Mabian Yi Autonomous county was 215,499 as of 2015 and has a total area of 2,304 square kilometers. Of the total area, a large area consists of mountains and rivers, and 87.3% consists of mountains with elevations higher than 1,000 meters. In 2015, GDP per capita was US$2,390 (14,888 CNY) in Mabian Yi Autonomous County whereas the province and national average were US$5,937 (36,981 CNY) and US$11,063 (68,905 CNY) respectively.[1] In addition, the disposable personal income was US$1,380 (8,595 CNY). With a poverty rate of 12.6%, it has been one of the 585 key counties in the development-oriented poverty reduction program for many years.

In Mabian Yi Autonomous County, 47.51% of the residents are the Yi people, and the ratio is even higher in some remote rural areas. Many of the Yi households do not value youth's education seriously as many families live far away from schools and expect children to help with farm chores and housework as much as possible. Furthermore, if the Ministry of Education did not implement the nine-year compulsory education, some parents would not plan to have their children attend secondary school after graduating from primary school. According to the survey conducted at the beginning of the fall 2015 semester, 20% of the students commuted more than one hour, one way to school. After school, students spent on average more than two hours on farm chores and housework. These conditions provided an environment to produce low-performing schools and students, and there is much room for the average scores of the countywide standardized examinations to be improved.

Table 3.1 presents the performance distribution of the countywide standardized test in Chinese and mathematics in the pre-treatment semesters (spring 2015 and spring 2016) for students in the sample. In the spring 2015 semester, the average grade of the countywide standardized test in Chinese was 57.95 out of 100 for the sixth-grade students in the sample, with only about 54.37%

[1] Data on the GDP for the year 2015 come from the Annual Report on Economics and Social Development of Mabian Yi Autonomous county and the National Bureau of Statistics of China.

of the students scoring higher than a passing score, 60. Mathematics performance was even lower with an average of 46.82, and there was only 38.29% of the students scoring higher than 60. Moreover, only 33.73% of the students scored higher than 60 in both subjects. In the spring 2016 semester, the third-grade students in the sample scored on average a 52.85 in Chinese, and only 47.67% of the students scored higher than 60. The average mathematics score was 57.16 for the same third-grade students, and 52.48% of them received a score higher than 60. Roughly 38.78% of the students scored higher than 60 in both subjects. For the fourth-grade students, the average Chinese score was 57.62 while 55.85% of the students scored higher than 60. The average mathematics score was 62.08, and 61.17 % of the fourth-graders obtained a score higher 60. 48.4% of the students scored higher than 60 in both subjects.

The importance of primary and middle schools is usually ignored as people usually lack information on the returns to primary and lower secondary education and irrationally discount the future benefits of obtaining it. With the relatively low level of GDP per capita in the county and the poor academic performance of the students across most of the school districts, appropriate monetary incentives tied to school performance is a useful tool that could be used to improve student outcomes. Additionally, a well-designed incentive program could be as an instrument to positive change both student and parent attitudes and behaviors toward educational activities. The rationale for utilizing conditional cash transfers is that by offering monetary incentives tied to multiple measures of performance, parents and students could invest more time and effort in educational services. The investment could in-turn improve student academic achievement, increase human capital accumulation, and decrease the possibility of inter-generational transmission of poverty. In addition, it is expected to be an ideal place to evaluate the impact of an incentive program on students' academic achievement.

3.2.1 Sample Design and Implementation

The YEIP was launched by the Research Center of China Household Finance Survey at Southwestern University of Finance and Economics in the fall 2015 semester. The program aimed at improving students' academic achievement through the conditional cash transfer in the short-run

Table 3.1: Pre-treatment Performance

Grade	N	Subject	Mean	5%	25%	50%	75%	95%	Pr(score >= 60)
		Chinese	57.95	14	39	63	77	89	54.37%
5^{th}(Spring 2015)	504	Mathematics	46.82	6	18	47	73	91	38.29%
		Total	104.77	25	63	109	149	175	33.73%[a]
		Chinese	52.85	4	27	57	81	93	47.67%
3^{th}(Spring 2016)	686	Mathematics	57.16	8	33	63	82	93	52.48%
		Total	110.02	14	64	123	160	180	38.78%
		Chinese	57.62	9	38	66	79	90	55.85%
4^{th}(Spring 2016)	753	Mathematics	62.08	10	40	69	87	96	61.17%
		Total	119.69	23	81	135	161	182	48.40%

Note. –

The countywide standardized examinations were out of 100 points in both subjects.

[a]This indicates the total percentage of students who scored higher than 60 in both Chinese and mathematics.

and overcoming poverty through the accumulation of human capital in the long-run. In the fall 2015 semester, randomization was performed using a class-school-based randomization design where all students in a class were in the same group. There were a total of 28 schools with 81 sixth-grade classes and 3,323 sixth-grade students in Mabian Yi Autonomous County in the 2015-2016 academic year. The 28 schools were first sorted by the average school score. Nine schools were picked up using probability proportional to size sampling method where schools across various performance could be included in the incentive program. Nine classes from the nine schools were randomly selected to be in the treatment group and were offered conditional cash transfers. Control classes were randomly selected from the nine schools and received no incentive. Classes were assigned to the control or treatment groups by the YEIP; they were not given a choice of being treated or not being treated.

In addition, another four classes were randomly chosen from four of the nine selected schools. In these four classes, students and teachers were both offered incentive payments in return for achieving multiple measures of performance. Given the truth that most of conditional cash transfer programs offer monetary incentives to only students or to teachers, the additional treatment group of both students and teachers receiving benefits could allow me to estimate the difference in the effects of the incentives between different experimental designs. Specifically, I could be able to measure the effectiveness of three alternative incentive schemes if the experiment was implemented consistently and steadily. However, this scheme was terminated soon after the conditional cash transfer program was implemented in Mabian Yi Autonomous County in 2015 due to teachers, whom were not offered any monetary incentives, discovering the program and issuing complaints about not being treated and receiving incentive payments. Therefore, I analyze only the effect of the treatment that applied only to students.

In the fall 2016 semester, the incentive program was repeated in the same county. There were 116 fourth- and 85 fifth-grade classes with 3,892 and 3,852 students from the 28 schools. The schools from the previous year and an additional five schools were selected. Most of the selected schools were randomly allocated to either control or treatment groups; however, three of them

consisted of both treatment and control groups. Among the 14 selected schools, nine schools had only control classes, two schools had only treatment classes, and three schools had both control and treatment groups. In the control schools, all fourth- and fifth-grade classes were assigned as control classes. All fourth- and fifth-grade classes were assigned as treatment classes in the treatment schools. In the schools with both the control and treatment groups, half of all the classes were assigned as control classes, and the other half were assigned as treatment classes.

3.2.2 Surveys

Surveys were first conducted to collect information on the individual, household, teacher, and class characteristics before treatment occurred each semester. Students were asked to answer questions regarding demographics, such as gender, ethnicity, and their study behavior. Moreover, questions designed to measure a student's intelligence quotient and emotional quotient were included at the end of the questionnaire to better understand students' academic ability, the capability of recognizing and adjusting their own emotions, and how they react to those of others. Students' families were also interviewed to collect information on household characteristics. Among the many variables from the household questionnaire, I utilize the parental education and household expenditure on the educationally related category. The data from the teacher and the school surveys include homework completion rates, teacher's time management, and student performance in the standardized examinations.

Part of the schools was not interviewed for individual and household surveys. Table 3.2 provides an insight into the sample attrition. Seven sixth-grade control classes and seven sixth-grade treatment classes from the seven schools in the fall 2015 semester were included in the sample for program analysis. In total, there were 241 sixth-grade students in the control group and 263 sixth-grade students in the treatment group. In 2016, 12 fourth-grade classes from the nine control schools and three fourth-grade classes from the two treatment schools were interviewed. In addition, one control class and five treatment classes from the schools with both control and treatment groups were interviewed. For the fifth-grade, the YEIP interviewed 12 classes from the nine control schools and three classes for the treatment schools. Two control classes and five treatment

classes from the schools with control and treatment groups were also interviewed. In 2016, 451 fourth-grade students and 466 fifth-grade students were included in the control group; and 235 fourth-grade students and 286 fifth-grade students were included in the treatment group.

3.2.3 Countywide Standardized Examinations

At the end of each semester, including the fall and spring semesters, the County Office of Education administers the countywide standardized examinations on Chinese and mathematics to students in each grade. The purpose of giving examinations at the end of each semester is to assess whether students have learned what they were expected to learn or and to what degree they have learned the material. Scores on examinations are not used as a requirement to transit from one grade to the next or from primary school to secondary school. The incentives were provided based on the performance of the examinations at the end of the fall semesters in 2015. In addition, performances on the countywide standardized Chinese and mathematics examinations administered in each primary school at the end of the pre-treatment semester were taken as the baseline grade to understand the effects of the incentives better. Examination data for individual students on the countywide standardized final exams were obtained from each school. Examination scores were normalized such that scores in the control group have a mean of 0 and a standard deviation of 1 within each grade in each semester.

3.2.4 Randomization Evidence

Table 3.3 presents the quality of the randomization by comparing the treatment and control classes based on pre-treatment characteristics and performance. The first category in part (A) shows student related characteristics, the proportion of Han Chinese, proportion of boys, mean intelligence quotient, and mean emotional quotient. For students in the fall 2015 and fall 2016 semester in the sample, all variables of the treatment group in this category did not differ from those of the control group at conventional levels of statistical significance. These results indicate that students from control and treatment groups did not differ in average individual characteristics. The second category in part (B) includes three household characteristics: the proportion of fathers

Table 3.2: Summary Sample Sizes

		6^{th}(Fall 2015)		4^{th}(Fall 2016)		5^{th}(Fall 2016)	
		Control	Treatment	Control	Treatment	Control	Treatment
	Total	28 (81 classes)		28 (116 classes)		28 (85 classes)	
School	Program	9 (36 classes)		9+3 (36+10 classes)	2+3 (9+9 classes)[a]	9+3 (34+9 classes)	2+3 (10+9 classes)[b]
	Analysis	7 (27 classes)		9+1 (36+2 classes)[c]	2+2 (9+6 classes)[d]	9+2 (34+5 classes)[e]	2+2 (10+7 classes)[f]
Class	Program	9	9	36+10	9+9	34+9	10+9
	Analysis	7	7	12+1	3+5	12+2	3+5
N	Analysis	241	263	451	235	466	286

[a] Three schools were chosen as the control and treatment at the same time.
[b] Three schools were chosen as the control and treatment at the same time.
[c] The control group in two of the three overlapping schools was not surveyed.
[d] The treatment group in one of the three was not surveyed.
[e] The control group in one of the three overlapping schools was not surveyed.
[f] The treatment group in one of the three was not surveyed.

and mothers with high school diploma and household educational expenditure. Parental education of treatment group did not depart from that of the control group at conventional levels of statistical significance for the fall 2015 sample. The average educational expenditure for the treatment and control groups were not different from each other at conventional levels of statistical significance for the fall 2015 sample either. These indicate that students in the control and treatment groups tended to come from families where parents have similar years of educational attainment and were willing to spend a similar amount on educational activities. Conversely, parental education of the treatment group departed from that of the control group at the 0.05 significance level for the students form the 2016 fall semester. The average educational expenditures for the treatment and control groups were different from each other at a 0.01 significance level for the students in the 2016 fall semester. The differences in the household features between the treatment and the control groups in the fall 2016 semester indicate that students in the treatment group tended to come from families where parents had significantly more years of educational attainment and were willing to spend more on educational activities. The last category provides a quick overview of some critical variables obtained from the homeroom teacher's response in the survey. All the variables of the treatment group on the class level did not differ from those of the control group at conventional levels of statistical significance for the fall 2015 sample. For the fall 2016 semester, the homework completion rate of the control group was lower than that of the treatment group. All other variables of the treatment group were not different from those of the control group at conventional levels of statistical significance.

Part (D) compares control and treatment groups on the basis of the average normalized scores on Chinese and mathematics in the baseline semester. The average scores in Chinese in the baseline semester did not differ between the control and treatment groups at conventional levels of significance (-0.10 versus -0.10 with p-value 0.99). Similarly, the average mathematics score of the treatment group in the baseline did not depart from that of the control group at conventional levels of significance (-0.09 versus -0.17 with p-value 0.85). In contrast, the Chinese score for the fourth- and fifth-grade students in the treatment group in the fall 2016 semester was higher than

Table 3.3: Comparison of Treatment and Control: Summary Characteristics

	6^{th} grade (Fall 2015)			4^{th}/5^{th} grade (Fall 2016)		
	Control	Treatment	p-value	Control	Treatment	p-value
Number of classes	7	7		27	16	
N	241	263		917	521	
Class size	34.42	37.57	0.50	33.96	32.56	0.78
(A) Variables from student survey:						
Han Chinese	0.29	0.34	0.82	0.20	0.32	0.21
Boy	0.53	0.52	0.85	0.52	0.53	0.81
Intelligence quotient	99.32	97.21	0.88	90.78	101.74	0.13
Emotional quotient	97.48	100.74	0.69	99.47	97.20	0.63
(B) Variables from household survey:						
Father with high school diploma	0.21	0.18	0.78	0.16	0.33	0.02
Mother with high school diploma	0.12	0.12	0.99	0.09	0.22	0.03
Educational expenditure	US$97.46 (607.00 CNY)	US$80.86 (503.59 CNY)	0.45	142.99 (890.52 CNY)	200.76 (1250.36 CNY)	0.03
(C) Variables from teacher survey:						
Homework completion rate	0.85	0.70	0.27	0.67	0.88	0.00
Working hours	41.71	37.71	0.34	30.36	27.65	0.60
Preparation hours	9.14	7.85	0.58	9.21	7.16	0.17
(D) Normalized scores on examinations at baseline:						
Chinese	-0.10	-0.10	0.99	-0.10	0.34	0.06
Mathematics	-0.09	-0.17	0.85	-0.10	0.33	0.07

Note. –
All examinations have been normalized within each grade in each semester such that scores in the control group are distributed with mean 0 and standard deviation of 1.

that in the control group (-0.10 versus 0.34 with p-value 0.06). The average score in mathematics for the fourth- and fifth-grade students in the treatment group in the fall 2016 semester was higher than that in the control group (-0.10 versus 0.33 with p-value 0.07).

3.2.5 The Incentive Program

Merit-based scholarships have been frequently criticized for the possible adverse equity effects that relatively advantaged students are the primary beneficiaries of these types of awards. The YEIP was a pilot experiment designed to improve youth's academic achievement extensively in neediest areas through various cash transfer channels within the program and thus accumulate human capital on the condition that pre-specified requirements were met. There were three different channels provided in the conditional cash transfer program. The first award offered cash incentives to students who achieved distinguished performance on examinations in Chinese and mathematics. Students were rewarded according to the rank in their classes. The higher the position, the larger the award amount. The second award rewarded students for examination progress. Students who received the top 10% of improvement were rewarded for their efforts. The third award offered incentive payments for outstanding homework achievement. The last two awards targeted students with relatively poor performance on examinations, who had a small probability of winning the first or the second awards. With YEIP, students could receive rewards by pursuing distinguished performance in examinations, improving examination scores, or achieving high homework grades. That means, not only relatively advantaged students could receive an award based on performance in examinations, but also students with little chance of winning a distinguished examination performance award could be encouraged by the program through other channels. I thus can examine the impacts of the conditional cash transfer program that considers multiple measures of performance including grades on the examinations, improvement on examinations, and homework grades.

For distinguished examination performance award, students received an incentive payment of US$48.2 (300 CNY) if they scored in the top 4% of the summed Chinese and mathematics examination scores in their class. Students received a payment of US$32.1 (200 CNY) if they scored in the top 4-8% in the class, and received US$24.1 (150 CNY) if they scored in the top 8-12% in

the class. The awards in this channel of the cash transfer program were received by mainly well-performed students. For examination progress award, students received a payment of US$24.1 (150 CNY) for reaching a top 10% of improvement in their class. This award was designed to encourage students who had little chance of winning the distinguished examination performance award and incorporate more students in the incentive program. For the outstanding homework award, an incentive payment of US$8.0 (50 CNY) was given to students who received a top 10% grade in their homework assignments during the semester. Many of the students who performed relatively poor and had never received an award before this program were eligible for an award in the last two categories. Homework assignments are believed to help students learn the material and prepare for the examinations and, hence, should impact examination grades as well. Tables 3.4 provides the detail of the incentive program. Students in the control group were not informed of the treatment throughout the semester. The awards were substantial considering that disposable income was only US$1,380 (8,595 CNY) for rural areas and most of the families in the remote rural area had an income far below the rural average.

Table 3.4: Incentives Design

Incentive Category	Criterion	Incentive Amount
Distinguished Examination Performance	Top 4% in class	US$48.2 (300 CNY)
	4-8% in class	US$32.1 (200 CNY)
	8-12% in class	US$24.1 (150 CNY)
Examination Progress	Top 10% of progress in class	US$24.1 (150 CNY)
Outstanding Homework	Top 10% in class	US$8.0 (50 CNY)

Tables 3.5 summarizes the awards rewarded in the treatment semesters (the fall 2015 and 2016 semesters). In the fall 2015 semester, 49 students were awarded for distinguished performance on

examinations. Additionally, 41 and 23 students received cash payments for improvement in examination performance and outstanding homework assignments respectively. In the fall 2016 semester, 73 fourth-grade students and 73 fifth-grade students received the distinguished examination performance reward. 59 fourth-grade students were rewarded for examination progress, and 77 students were rewarded for outstanding homework. 59 fifth-graders won awards for examination progress, and 75 students received awards for outstanding homework.

Table 3.5: Number of Awards Granted

		6^{th}(Fall 2015)		4^{th}(Fall 2016)		5^{th}(Fall 2016)	
Incentive	Amount	N	Subtotal	N	Subtotal	N	Subtotal
Top 4% in class	300 CNY	12	3,600	18	5,400	16	4,800
Top 4-8% in class	200 CNY	22	4,400	22	4,400	27	5,400
Top 8-12% in class	150 CNY	15	2,250	33	4,950	30	4,500
Examination Progress	150 CNY	41	6,150	59	8,850	59	8,850
Outstanding Homework	150 CNY	23	920	77	3,850	75	3,750
Total			17,320		27,450		27,300

3.3 Empirical Results

In this essay, I focus on the reduced-form estimation of the program effect on examination scores given scores received in the previous semester. The estimation equation is

$$E_{i,c,k,j,t} = \alpha + \beta_1 E_{i,c,k,j,t-1} + \gamma T_{c,k,j,t} + X'_{i,c,k,j,t}\zeta + \text{ grade}_k + \text{ school}_j + \epsilon_{i,c,k,j,t}. \tag{3.1}$$

for $i = 1, 2, \ldots, n$, $c = 1, 2, \ldots, C$, $k = 1$, $j = 1, 2, \ldots J$, and $t = 1, 2, \ldots, T$. $E_{i,c,k,j,t}$ is the normalized examination score in either Chinese or mathematics for student i in class c grade k at school j in the program semester, i.e. the fall 2015 and fall 2016 semesters, and $E_{i,c,k,j,t-1}$ is the normalized score for student i in class c grade k at school j in the baseline semester, i.e. the spring 2015 and spring 2016 semesters. $T_{c,k,j,t}$ is the treatment class indicator, and the coefficient γ captures the average treatment effects on the student who was offered incentive payments based on

multiple measures of performance. $X'_{i,c,k,j,t}$ is a vector that includes the individual, household, and class controls. school_j is the school fixed effect indicator. grade_k is the grade fixed effect indicator. The variable grade is included when I analyze the effects of conditional cash transfer program for the fall 2016 semester since both fourth- and fifth-grade students were offered incentive payments if specific conditions were met ($\text{grade}_k = 1$ if students came from fifth-grade in the fall 2016 semester). The error term $\epsilon_{i,c,k,j,t}$ captures unobserved individual characteristics or idiosyncratic shocks.

Furthermore, I am interested in the heterogeneous treatment effects among students with different baseline performance. I add interaction terms of the baseline performance quartile and the incentive treatment to evaluate the heterogeneous effects among students with various baseline performance. I estimate the following equation:

$$\begin{aligned} E_{i,c,k,j,t} =& \alpha' + \beta'_{12}\text{Third quartile of } E_{i,c,k,j,t-1} + \beta'_{13}\text{Second quartile of } E_{i,c,k,j,t-1} \\ &+ \beta'_{14}\text{Top quartile of } E_{i,c,k,j,t-1} + \gamma'_{11}\text{Bottom quartile of } E_{i,c,k,j,t-1} \times T_{c,k,j,t} \\ &+ \gamma'_{12}\text{Third quartile of } E_{i,c,k,j,t-1} \times T_{c,k,j,t} + \gamma'_{13}\text{Second quartile of } E_{i,c,k,j,t-1} \times T_{c,k,j,t} \\ &+ \gamma'_{14}\text{Top quartile of } E_{i,c,k,j,t-1} \times T_{c,k,j,t} + X'_{i,c,k,j,t}\zeta' + \text{ grade}_k + \text{ school}_j + \epsilon_{i,c,k,j,t}. \end{aligned} \tag{3.2}$$

where the bottom, third, second, and top quartiles of $E_{i,c,k,j,t-1}$ indexes the quartile of achievement for students in pre-treatment test scores. Other definitions of variables for equation (3.2) are the same with those for equation (3.1).

3.3.1 Treatment Effects on Chinese Examination

Table 3.6 reports the regression estimates of regression (3.1) when the dependent variable is the normalized score in Chinese received at the end of the program semester, i.e., the fall 2015 and fall 2016 semesters. Student individual, household, and class features are listed in Table 3.3 and are all included in the regression analysis. The first two columns show the coefficient estimates for equations (3.1) and (3.2) for the sixth-grade students in the fall 2015 semester. The baseline well

predicted the score in the treatment semester, and the incentive significantly raised score in Chinese by 0.133 standard deviations. I further examine how the incentive affected scores heterogeneously according to the baseline score in the previous semester by estimating equation (3.2). The estimates are shown in the second column. I find that the incentives had demonstrated a positive effect on the score especially, particularly for students with a bottom quartile baseline performance with an estimate of 0.367 standard deviations. For students with a baseline score that was in the third, second, and top quartiles, the impact of the incentives was not deterministic or significant.

In the third column, I find that the baseline score played an essential role in determining the score in the program semester. Differing than the fall 2015 sample, the estimates show that the incentives did not increase the score for the fourth- and fifth-grade students in the fall 2016 semester. I also investigate if the incentives affected scores heterogeneously according to the baseline score in the previous semester in the fourth column. Although the incentives demonstrated a positive effect on the score for students in the bottom quartile of baseline performance (0.134 standard deviations), large standard errors produce an insignificant effect.

3.3.2 Treatment Effects on Mathematics Examination

I now discuss regression analysis of the impacts of the incentives on student performance in Mathematics in Table 3.7. Similarly, baseline mathematics score was a strong predictor of mathematics score in the program semester for students in both the fall 2015 and fall 2016 samples, see column 1 and column 3. The first column shows that the treatment increased score in mathematics by 0.306 standard deviations for the six-grade students in the fall 2015 sample. Furthermore, the second column shows that there was a significant positive impact of 0.525 standard deviations for students a bottom quartile baseline score and a smaller effect of 0.170 standard deviations for students with a third quartile baseline score. The effects of the incentives on the score for students with second and top quartile baseline performance were not significant.

I then examine the effect of the treatment on the score for the fourth- and fifth-grade students in the fall 2016 sample. The impact was negative but not significant. I again investigate the possible heterogeneous program impact on student performance. Although the estimates were positive

Table 3.6: Treatment Effects on Examination Scores: Chinese

	Dependent variable: normalized score			
	6^{th} grade (Fall 2015)		4^{th}/5^{th} grade (Fall 2016)	
	(1)	(2)	(1)	(2)
Incentive treatment	0.133		0.016	
	(0.059)**		(0.163)	
Baseline score	0.688		0.818	
	(0.049)***		(0.016)***	
Incentive treatment × baseline score in bottom quartile		0.367		0.134
		(0.097)***		(0.157)
Incentive treatment × baseline score in third quartile		0.065		-0.096
		(0.101)		(0.178)
Incentive treatment × baseline score in second quartile		-0.080		-0.127
		(0.101)		(0.185)
Incentive treatment × baseline score in top quartile		-0.002		-0.168
		(0.159)		(0.185)
Individual/ household/ class characteristics	Yes	Yes	Yes	Yes
âĂćR^2	0.84	0.82	0.85	0.81
N	504	504	1,438	1,438

Note. –
* Significant at 10%. ** Significant at 5%. *** Significant at 1%.
All examinations have been normalized within each grade in each semester such that scores in the control group are distributed with mean 0 and standard deviation of 1.

Standard deviations in parentheses were constructed by drawing 300 bootstrap replications.

for students with lower baseline scores and negative for students with higher baseline scores, the effects were insignificant throughout all different baseline quartiles.

3.4 Conclusion

I evaluate the impact of the YEIP implemented in Mabian Yi Autonomous County. Mabian Yi Autonomous County is a relatively poor county located in southwest China's Sichuan province. The estimates show the desired results regarding the impact of the incentive program for the sample with balanced characteristics. In contrast to the results found by Bettinger (2012), students who already performed well were most incentivized by the pay-for-performance program, the cash incentive program in Mabian Yi Autonomous County demonstrated the most significant effect for students with relatively poor baseline performance. For the fall 2016 sample in which characteristics were not balanced, I do not observe an impact of the conditional cash transfer program on any test scores.

I am currently in the process of obtaining the post-treatment academic performance for students in my analysis. Using the post-treatment performance data, I plan to explore the longer-term impacts of the conditional cash program by estimating the score differences in junior high school after students graduated from the primary school between students in the control and treatment groups. Results from these data will provide additional insight into the full range of impacts conditional cash transfers have in shaping child educational decision and outcomes.

Table 3.7: Treatment Effects on Examination Scores: Mathematics

	Dependent variable: normalized score			
	6^{th} grade (Fall 2015)		4^{th}/5^{th} grade (Fall 2016)	
	(1)	(2)	(1)	(2)
Incentive treatment	0.306		-0.023	
	(0.065)***		(0.167)	
Baseline score	0.627		0.769	
	(0.029)***		(0.020)***	
Incentive treatment × baseline score in bottom quartile		0.525		0.001
		(0.103)***		(0.180)
Incentive treatment × baseline score in third quartile		0.170		0.129
		(0.093)*		(0.181)
Incentive treatment × baseline score in second quartile		-0.034		-0.204
		(0.118)		(0.189)
Incentive treatment × baseline score in top quartile		0.137		-0.202
		(0.121)		(0.185)
Individual/ household/ class characteristics	Yes	Yes	Yes	Yes
R^2	0.80	0.79	0.80	0.76
N	504	504	1,438	1,438

Note. –
* Significant at 10%. ** Significant at 5%. *** Significant at 1%.
All examinations have been normalized within each grade in each semester such that scores in the control group are distributed with mean 0 and standard deviation of 1.
Standard deviations in parentheses were constructed by drawing 300 bootstrap replications.

4. CONCLUSIONS

This dissertation includes two independent essays. In the first essay, I employ a selection bias correction and equivalence scale method to estimate the costs of raising children that consider economies of scale enjoyed by a couple living with children. The second essay evaluates the impact of a social experiment that designed to improve youth's academic achievement in a neediest area through conditional cash transfers. Below I summarize these two essays.

The Chinese government replaced the 35-year-old one-child policy with a two-child policy on January 1, 2016. A number of studies have been interested in the impact of the recent change in the family planning policy in China. The new policy raised births by 1.31 million in China in 2016 shortly. Births then decreased by 0.63 million in 2017. The government originally estimated the new policy would increase 3 million newborns annually in the five or six years following the policy change. The increase in 2016 fell far short of the government's target and could not compensate the decrease in the willingness of having a first child. According to fertility intention and behavior surveys conducted in China, the low willingness of having a child or having an additional child was attributed to the high costs of raising children. Cost estimates of raising children can provide an important insight into how children affect family economic decisions. In China, couples with two children are over-represented by a low-educated, rural, and sons-preferred population. This is commonly attributed to an exemption in the one-child policy: only particular couples could have a second child, and these couples were disadvantaged, lived in rural areas, and had a strong preference for sons. Using data from the CHFS conducted in 2011, I employ a selection bias correction and equivalence scale method to estimate the costs of raising children that consider economies of scale enjoyed by a couple living with children. I find no difference between the costs of raising a son and a daughter for a representative one-child family. The marginal cost of raising an additional child was higher for a representative couple with a firstborn daughter and a second-born son than a couple with children of the same gender. To raise an additional child, on average, the cost was higher for an urban family than a rural one, and the cost was higher for a couple with

more years of schooling than one with fewer years of education.

Besides the family planning policy, another subject-matter is China's unequal access to educational resources among children from different backgrounds. Many countries have introduced conditional cash transfer programs, which provide incentives to relatively disadvantaged individuals or households when specific requirements are met. I evaluate the impact of the YEIP that designed to improve youth's academic achievement in a neediest area through conditional cash transfers. The estimates show that the incentive had positive impacts on student performance in both standardized Chinese and mathematics examinations when using a sample with balanced characteristics. I also estimate the impacts by baseline performance and find the most significant effects of the monetary incentives for students with a relatively low baseline score while finding no significant impacts for students with a relatively high baseline score. Results using a sample with unbalanced characteristics yield no significant impacts across different baseline scores.

REFERENCES

Angrist, Joshua and Victor Lavy, "The Effects of High Stakes High School Achievement Awards: Evidence from a Eandomized Trial," *American Economic Review*, 2009, *99* (4), 1384–1414.

Bargain, Olivier and Olivier Donni, "Expenditure on Children: A Rothbarth-type Method Consistent with Scale Economies and Parents' Bargaining," *European Economic Review*, 2012, *56* (4), 792–813.

Barrow, Lisa and Cecilia E. Rouse, "Financial Incentives and Educational Investment : The Impact of Performance-based Scholarships on Student Time Use," *National Bureau of Economic Research Working Paper 19352*, 2013.

Becker, Gary S., "An Economic Analysis of Fertility," in "Demographic and Economic Change in Developed Countries," Princeton, NJ: Columbia University Press, 1960, pp. 209–240.

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin, "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment In Mexican High Schools," *Journal of Political Economy*, 2015, *123* (2), 325–364.

Bettinger, Eric P., "Paying to Learn: The Effect of Financial Incentives on Elemantary School Test Scores," *Review of Economics and Statistics*, 2012, *94* (3), 686–698.

Browning, Martin, "Children and Household Economic Behavior," *Journal of Economic Literature*, 1992, *30* (3), 1434–175.

_ , Pierre-André Chiappori, and Arthur Lewbel, "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *Review of Economic Studies*, 2013, *80* (4), 1267–1303.

Buttner, Thomas and Wolfgang Lutz, "Estimating Fertility Responses to Policy Measures in the German Democratic Republic," *Population and Development Review*, 1990, *16* (3), 539–555.

Cohen, Alma, Rajeev Dehejia, and Dimitri Romanov, "Financial Incentives and Fertility," *Review of Economics and Statistics*, 2013, *95* (1), 1–20.

de Brauw, Alan and John Hoddinott, "Must Conditional cash Transfer Programs Be Conditioned to Be Effective? The impact of Conditioning Transfers on School Enrollment in Mexico," *Journal of Development Economics*, 2011, *96* (2), 359–370.

_ **, Daniel O. Gilligan, John Hoddinott, and Shalini Roy**, "The Impact of Bolsa Família on Schooling," *World Development*, 2015, *70*, 303–316.

Deaton, Angus S., Javier Ruiz-Castillo, and Duncan Thomas, "The Influence of Household Composition on Household Expenditure Patterns: Theory and Spanish Evidence," *Journal of Political Economy*, 1989, *97* (1), 179–200.

Dunbar, Geoffrey R., Arthur Lewbel, and Krishna Pendakur, "Children's Resources in Collective Households: Identification, Estimation, and an Application to Child Poverty in Malawi," *American Economic Review*, 2013, *103* (1), 438–471.

Engel, Ernst, "Die Lebenskosten Belgischer Arbeiter-Familien Fruher und Jetzt," *Bulletin de L'Institut International de Statistique*, 1895, *9* (1), 1–124.

Fryer, Roland G., "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *Quarterly Journal of Economics*, 2011, *126* (4), 1755–1798.

Glewwe, Paul and Ana Lucia Kassouf, "The Impact of the Bolsa Escola/Familia Conditional Cash Transfer Program on Enrollment, Dropout Rates and Grade Promotion in Brazil," *Journal of Development Economics*, 2012, *97* (2), 505–517.

_ **, Nauman Ilias, and Michael Kremer**, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2003, *2* (3), 205–227.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel, "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 2011, *25* (4), 191–210.

Gronau, Reuben, "Consumption Technology and the Intrafamily Distribution of Resources: Adult Equivalence Scales Reexamined," *Journal of Political Economy*, 1988, *96* (6), 1183–1205.

_ , "The Intrafamily Allocation of Goods–How to Separate the Adult from the Child," *Journal of Labor Economics*, 1991, *9* (3), 207–235.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47*

(1), 153–161.

Hoem, Jan M., "Social Policy and Recent Fertility Change in Sweden," *Population and Development Review*, 1990, *16* (4), 735–748.

Horvitz, Daniel G. and Donovan J. Thompson, "A Generalization of Sampling Without Replacement From a Finite Universe," *Neuropsychologia*, 1952, *47* (260), 663–685.

Hyatt, Douglas E. and William J. Milne, "Can Public Policy Affect Fertility?," *Canadian Public Policy 17:1*, 1991, *17* (1), 77–85.

Kornrich, Sabino and Frank Furstenberg, "Investing in Children: Changes in Parental Spending on Children, 1972-2007," *Demography*, 2013, *50* (1), 1–23.

Kremer, Michael, Edward Miguel, and Rebecca Thornton, "Incentives to Learn," *Review of Economics and Statistics*, 2009, *91*, 437–456.

Laroque, Guy and Bernard Salanié, "Identifying the Response of Fertility to Financila Incentives," *Journal of Applied Econometics*, 2014, *29*, 314–332.

Levitt, Steven, John A. List, and Sally Sadoff, "The Effect of Performance-Based Incentives on Educational Achievement: Evidence From a Randomized Experiment," *National Bureau of Economic Research Working Paper 22107*, 2016.

Lewbel, Arthur and Krishna Pendakur, "Estimation of Collective Household Models with Engel Curves," *Journal of Econometrics*, 2008, *147* (2), 350–358.

Lindahl, Erik, "Just Taxation - A Positive Solution," in Richard Musgrave and Alan T. Peacock, eds., *1958, Classics in the Theory of Public Finance*, London: MacMillan, 1919.

Lino, Mark, Kevin Kuczynski, Nestor Rodriguez, and Tusarebecca Schap, "Expenditures on Children by Families , 2015," Technical Report, United States Department of Agriculture Center 2017.

List, John A., Jeffrey A. Livingston, and Susanne Neckermann, "Do Financial Incentives Crowd Out Intrinsic Motivation to Perform on Standardized Tests?," *Economics of Education Review*, 2018, *66*, 125–136.

Manski, Charles F. and Joram Mayshar, "Private Incentives and Social Interactions: Fertility

Puzzles in Israel," *Journal of the European Economic Association*, 2003, *1* (1), 181–211.

Mao, Zhouyan, "An Empirical Analysis on the Discrepancy between FertilityWillingness and Behavior," *Population and Economics*, 2009, *2*, 16–22.

_ **and Luo Hao**, "Difference between Fertility Intention and Fertility Behavior for Women Subject to the Two-child Policy: An Empirical Study Based on the Theory of Planned Behavior," *Population Research*, 2013, *37* (1), 84–93.

Milligan, Kevin, "Subsidizing the Stork: New Evidence on Tax Incentives and Fertility," *Review of Economics and Statistics*, 2005, *87* (3), 539–555.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher {Performance} {Pay}: {Experimental} {Evidence} from {India}," *Journal of Political Economy*, 2011, *119* (1), 39–77.

Nelson, Julie A., "Household Equivalence Scales: Theory versus Policy?," *Journal of Labor Economics*, 1993, *11* (3), 471–493.

Rothbarth, Erwin, "Note on a Method of Determining Equivalent Incomes for Families of Different Composition," in Charles Madge, ed., *Wartime Patterns of Saving and Spending*, Cambridge: Cambridge University Press, 1943, chapter Appendix 4, pp. 123–130.

Schultz, T. Paul, "Impact of Progresa on School Attendance Rates in the Sampled Population," Technical Report, International Food Policy Research Institute, Washington, D.C. 2000.

_ , "The Impact of Progresa on School Enrollments," Technical Report, International Food Policy Research Institute, Washington, D.C. 2000.

_ , "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program," *Journal of Development Economics*, 2004, *74* (1), 199–250.

Skoufias, Emmanuel, "Progresa and Its Impacts on the Welfare of Rural Households in Mexico," Technical Report, International Food Policy Research Institute, Washington, D.C. 2001.

Visaria, Sujata, Rajeev Dehejia, Melody M. Chao, and Anirban Mukhopadhyay, "Unintended Consequences of Rewards for Student Attendance: Results from a Field Experiment in Indian Classrooms," *Economics of Education Review*, 2016, *54*, 173–184.

Walker, James R., "The Effect of Public Policies on Recent Swedish Fertility," *Journal of Popu-*

lation Economics, 1995, *8*, 223–251.

Whittington, Leslie A., "Taxes and the Family: The Impact of the Tax Exemption for Dependents on Marital Fertility," *Demography*, 1992, *29* (2), 215–226.

_ , "State income tax policy and family size: fertility and the dependency exemption.," *Public finance quarterly*, 1993, *21* (4), 378–398.

_ **, James Alm, and H. Elizabeth Peters**, "Fertility and the Personal Exemption : Implicit Pronatalist Policy in the United States," *American Economic Review*, 1990, *80* (3), 545–556.

Willis, Robert J., "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 1973, *81* (2), 14–64.

Zhang, Junsen, Jason Quan, and Peter Van Meerbergen, "The Effect of Tax-Transfer Policies on Fertility in Canada, 1921-1988," *Journal of Human Resources*, 1994, *29* (1), 181–201.

Zhang, Yinfeng, "An Empirical Study on the Cause and Effect of Discrepancy between Fertility Willingness and Behavior, 2015," Technical Report, Tianjin Academy od Social Sciences 2016.