ESSAYS ON CHINA'S HOUSEHOLD SAVING RATE AND EDUCATION POLICY

A Dissertation

by

YONGZHI SUN

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,	Li Gan
Co-Chair of Committee,	Yuzhe Zhang
Committee Members,	Yonghong An
	Jason Lindo
	Ren Mu
Head of Department,	Timothy Gronberg

May 2018

Major Subject: Economics

ABSTRACT

This essay contains three essays in applied microeconomics. The first and second paper study the China's household saving rate and the third studies the economic policy in China.

In Volume 119 of the Journal of Political Economy, a paper uses data from the Chinese Household Income Project (CHIP) 2002 and finds that increasingly unbalanced premarital sex ratio raise household saving rate of son-families and the rapid increase in premarital sex ratio can potentially explain about half of China's household saving rate increasing during 1990-2007. This paper reexamines the competitive saving motive. We first use local sex ratio inferred from 2000 China population census and same dataset CHIP 2002 to find the competitive saving motive only holds for the household in rich counties. We then use data from the China Household Finance Survey (CHFS) to show that competitive holds for the rural sample. The cross-regional evidence indicates that the competitive saving motive exists, but only in the rural area. By estimation and computation, an increase in sex ratio from 1985 to 2015 can explain about 28% of the actual increase of the increase of rural saving rate.

The second paper studies the role of income inequality interacting with liquidity constraints in explaining the high household saving rate in China. The predictions implied by a simple lifecycle heterogeneous agent model are consistent with data facts. Using three large nationally representative data sets, China Household Finance Survey (CHFS), China Family Panel Studies (CFPS), and Chinese Household Income Project (CHIP), we find robust evidence that (1) the rich save more; (2) the poor are more likely to face liquidity constraints, and the effect of liquidity constraints on household saving rate is significantly positive; (3) income inequality has a significant positive effect on aggregate household saving rate; and (4) the marginal propensity to consume out of transitory income for poor households is significantly higher than for rich households. Our study provides a policy implication that economic policy of reducing income inequality would lower the aggregate saving rate and

thus become a policy of economic transition and growth.

The third paper estimates the effect of the "Program of College Admission for Poor Counties" on high school education using data from 86 counties of Gansu province in north-western China. Applying a difference-in-differences approach, we show that the program significantly increases senior high school entrants by 99-224, and enrollments by 317-586 in per 100,000 population in the poor counties in Gansu after the policy started in 2012. Using the alternative measurement of outcomes, we show that it significantly increases entry rate by 1.3-7.6%, and enrollment rate by 1.2-7.3%. The results are robust to alternative model specifications and outcome measurements. Our findings indicate that this admission policy, which is motivated by addressing unequal access to college, effectively improves schooling at the high school level.

DEDICATION

To My Beloved Parents, Grandparents, and Parents-in-Law, Grandma-in-Law,

for your unconditional and endless love, support, and encouragement throughout my life.

To My Loving Wife Qiong Ma,

for your love, company, support, and dedication, which has been all inspiration for me.

ACKNOWLEDGMENTS

First, I would like to thank and be deeply indebted to Professor Gan for his always persistent and dedicated guidance, support, encouragement, and recommendation on my academic research and career goal. In this precious journey of continuous research and papers, he is much more than just a great academic advisor. Your word to me is an all-time inspiration and tremendous benefits on the past student journey and for the rest of my life. For dear Professors Gan, thank you very much, for everything.

Second, I would like to thank Professor Mu, Professor Zhang, Professor An, Professor Lindo. What I have learned from them is more than I could ever give them credit for here. They will serve excellent models for me in the upcoming academic life. I would like also to thank Professor Chen and Professor Wu in Fudan University, who lead me to the fantastic journey of economics study and let me know what a real teacher is like from the very beginning. They remain my go-to guy to talk about ideas and life. For all dear professors, thank you very important for providing personal and professional guidance in both research and life. The imprint for you will be everlasting.

Many thanks also go to my dear friends in this Aggie Land. You all made an enjoyable, exciting and unforgettable time of working, studying and living in the U.S. What a precious memory, forever and always for me!

Last, but most importantly, I would like to thank my parents, Yizong Sun and Ying Chi, grandparents, Renjun Sun, Wenhua Li, Xianquan Chi and Yusheng Ding, and parents-in-law, Xinde Ma and Baoping Sun, and grandparents-in-law, Hongying Peng for your unconditional love, endless support and infinite encouragement throughout my life. I am eternally grateful along my entire journey. And most of all, I wish to thank my loving wife, Qiong Ma, especially for your love, company, support, and dedication, I would never have been the person I am today without you.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Li Gan [chair] and Professor Yuzhe Zhang[co-chair] and Professor Yonghong An and Professor Jason Lindo of the Department of Economics and Professor Ren Mu of the Bush School of Government and Public Service.

The analyses depicted in Chapter 2 and 3 were conducted in part by Naibao Zhao and Liwei Yin of the Department of Economics; the analyses depicted in Chapter 4 was conducted in part by Song Zhou of the Department of Agricultural Economics. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a Graduate Assistantship from the Department of Economics at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. GENERAL INTRODUCTION

The main focus of my doctoral research is to apply causal inference technique to identify the effects of unbalanced local sex ratio and income distribution, liquidity constraint on saving rate of the household sector. In addition, also applying causal inference methodology to evaluate and identify the effect of the "Program of College Admission for Poor Counties" on high school education in current China.

In the first essay, we answer this question "Can the sex ratio explain high China's household saving rate today?". The high aggregate household saving rate is one of the unique features of the Chinese economy. Over the period of China rapid income growth, China's sex ratio, which is defined as the ratio of the number of boys at birth to the number of girls at birth, has also been experienced a rapid a growth. The common trend has captured economists' attention on linking the saving rate to the sex ratio. Can the sex ratio explain high China's household saving rate? Starting with Wei and Zhang (2011a), they argue a explanation called "competitive saving motive" that increasingly unbalanced premarital sex ratio requires both rural and urban son-families to save increasingly more to compete in the marriage market and show that the rapid increase in premarital sex ratio during 1990-2007 can potentially explain about half of the sharp rise in China's household saving rate.

In this paper, we reexamine the impact of the sex ratio on the household saving rate. First, we use the CHIP data — the same sources of data as used by Wei and Zhang (2011a) — in an attempt to replicate and extend the estimates reported in Wei and Zhang (2011a). We use local sex ratio inferred from 2000 China population census and find the competitive saving motive only holds for the household in rich counties. The effects vanish for CHIP 2002 full sample with census data in 2000, especially for poor households and households in poor counties. We argue that unbalanced migration of premarital male and female is the major explanation for the vanishing effect.

Next, we turn to another nationally representative data set China Household Finance

Survey (CHFS) 2015 wave for estimating the effect of the sex ratio on the household saving rate. The result indicates that the competitive saving motive still holds, but only in the rural area. Specifically, in the rural, the effects on households with a son is much smaller, and the effects on households with a daughter are also significant positive. However, there are no effects on urban households.

Finally, we use China population census and provincial panel data as the same data sources as Wei and Zhang (2011a) used, and adopt a more precise algorithm to compute the sex ratios of the 31 provinces for 7 years (1985, 1990, 1995, 2000, 2005, 2010, and 2015). Our results report that the competitive saving motive exists, but only in the rural area. By estimation and computation, an increase in sex ratio from 1985 to 2015 can explain about 28% of the actual increase of the increase of rural saving rate.

In the second essay, we show in this paper the role of income inequality interacting with liquidity constraints in explaining the high household saving rate in China. In a simple two-period model, households are heterogeneous in income and subjective discount factor, and whether the liquidity constraint is binding, consumption and saving rate are endogenously determined. The model generates several predictions consistent data facts: (1) the rich save more; (2) the proportion of constrained households for the poor is higher than that for the rich; (3) liquidity constraints would increase household saving rate. (4) when income inequality increases, the rich save even more, in the meanwhile, the poor would also save more due to the binding liquidity constraints, and thus the aggregate household saving rate would rise.

Using three sources of large, nationally representative household survey data, the China Household Finance Survey (CHFS), the China Family Panel Studies (CFPS), and the Chinese Household Income Project (CHIP), we provide direct empirical evidence implied by the theoretical model. We find that in China, (1) the top 20 percent permanent income householdsâĂŹ saving rate is 19–23 percent signiïňĄcant higher than the bottom 20 percent householdsâĂŹ. (2) the bottom 20 percent permanent income households are more likely to face a borrowing constraint, with a 12–20 percent signiïňĄcant higher probability; (3) the existence of liquidity constraints would lead to a significant increase of more than 20 percent in the household saving rate; (4) income inequality would have a significant positive impact on the aggregate household saving rate ath the county level, with a 1 point on a scale of 100 measure increase in the Gini coefficient leading to a rise of 0.2 percent in the aggregate saving rate; (5) the estimated MPC for the top 20 percent households range from 200 to 400 RMB per 1000 RMB, while for the bottom 20 percent households, the range from 600 to 900 RMB per 1000 RMB.

These findings would have important policy implications. The Chinese governmentâĂŹs policies on reducing the saving rate have not yet produced substantial results. If income inequality and liquidity constraints were the key reasons for the high aggregate household saving rate, the resulting policy would be drastically different. For example, it is appropriate for the Chinese government to design some income redistribution programs (such as EITC) to reduce income inequality or devote more resources to support the credit market development. An economic policy of tackling income inequality would lower the aggregate saving rate, thus becoming a policy of economic transition and growth.

The third essay uses difference-in-differences strategy to estimate the effect of a college admission policy on the senior high school education. This policy, the "Program of College Admission for Poor Counties", assigns exclusive admission quotas to enroll students from poor counties. We find that schooling at senior high school stage is effectively improved in terms of entry and enrollment. Our findings also have significant policy implications. This policy for addressing unequal access to college education in fact enhanced schooling at high school stage, which exhibits a "general equilibrium effect", as senior high school and college are connected in educational continuation decisions. This finding indicates that the policy effect meets one of the policy objectives, that is, it provides positive incentives for promoting pre-tertiary education.

2. COMPETITIVE SAVING HYPOTHESIS REVISITED

2.1 Introduction

In the second chapter, we revisit the "competitive saving motive" and answer this question "Can the sex ratio explain high China's household saving rate today?". The high aggregate household saving rate is one of the unique features of the Chinese economy. According to National Accounts of OECD Countries, China's household savings as a percentage of household disposable income have been one of the highest in the world. The household saving rate has been increasing steadily from 28 percent in 2000 to 37 percent in 2015.[1] Over the period of China rapid income growth, China's sex ratio, which is defined as the ratio of the number of boys at birth to the number of girls at birth, has also been experienced a rapid a growth, from 1.07 in 1982 to 1.15 in 2015.[2] The common trend has captured economists' attention on linking the saving rate to the sex ratio.[3]

Can the sex ratio explain high China's household saving rate? The first study that links the two variables together in explaining the "Chinese saving puzzle" is Wei and Zhang (2011a).[4,5] They argue an explanation called "competitive saving motive" that increasingly unbalanced premarital sex ratio require both rural and urban son-families to save increasingly more to ensure the success in an increasingly competitive marriage market and increase the saving rate of the household factor. Wei and Zhang (2011a) use data from the Chinese Household Income Project (CHIP) data and provincial-level panel data in China to show that the rapid increase in premarital sex ratio during 1990-2007 can potentially explain about

[1] See panel (1) of Figure A.1 in Appendix.

[2] See panel (2) of Figure A.1 in Appendix

[3] See Figure A.2 in Appendix.

[4] Modigliani and Cao (2004) refers to high China's household saving rate the "Chinese saving puzzle".

[5] To date, other explanations in the literature includes demographic changes (Modigliani and Cao, 2004; Horioka and Wan, 2007; Curtis, Lugauer, and Mark, 2015, etc), precautionary saving motive (Meng, 2003; Chamon and Prasad, 2010; He, Huang, Liu, and Zhu, 2017, etc), habit formation (Horioka and Wan, 2007; Carroll and Weil, 1994). Since our paper focuses on reconsidering the gender imbalance and the competitive saving motive, starting with Wei and Zhang (2011a), in explaining Chinese household saving rate, we do not expand space to review other important explanations and leave them in a review paper by Yang, Zhang, and Zhou (2012).

half of the sharp increase in China's household saving rate.

Why competitive saving motive needs this replication In fact, birth and premarital sex ratio imbalance both rural and urban is still important in current China. Migration for working and schooling is another historical process for the youth during the rapid industrialization and urbanization in China. The pattern of unbalanced sex ratio for rural and urban can be reshaped during the intra-provincial and inter-provincial migration along with urbanization process over the last decades with Chinese rapid economic growth. The saving rate of Chinese household sector has been increasing steadily from 28 percent to 37 percent since Wei and Zhang (2011a). Does the competitive saving motive of premarital aged son-families still hold in China today? Or the competitive effect is still a factor for some part of China? Because of the wide citation around the world and important implications for other Asian countries experiencing the same unbalanced premarital sex ratio other than China, Horioka and Terada-Hagiwara (2017) revisited the competitive saving motive and find a significant impact on the household saving rate using long-term time series data for 1975–2010 period in both India and Korea. Many countries, including India, Korea, Singapore, Vietnam, Taiwan, in fact, have also exhibited unbalanced sex ratios in the premarital age cohorts. Accordingly, the competitive saving motive may have played a quantitatively important role in the evolution of these countries' savings rates as well. In addition, though it is not as easy to estimate its effect, the competitive saving motive can still be present and important in countries with a balanced sex ratio. Therefore, in this paper, we still focus on China's premarital sex ratio long-term pattern and reexamine the impact of local sex ratio on the household saving rate.

First, we use the CHIP data — the same sources of data as used by Wei and Zhang (2011a) — in an attempt to replicate and extend the estimates reported in Wei and Zhang (2011a). We first show that replication of Wei and Zhang (2011a) and analysis based on CHIP 2002 along with 1990 China population census. We also do the extensive robustness check with other county-level characteristics. Moreover, we use local sex ratio inferred from 2000 China population census and find the competitive saving motive only holds for the household in rich counties. The effects vanish for CHIP 2002 full sample with census data in 2000, especially for poor households and households in poor counties. We argue that unbalanced migration of premarital male and female is the major explanation for the vanishing effect.

Next, we turn to another nationally representative dataset for estimating the effect of the sex ratio on the household saving rate. In particular, we use the China Household Finance Survey (CHFS) 2015 wave, which covers a larger sample of individual and households and contains more detailed information about income, assets, debts, and expenditures. Estimated effects of the sex ratio on household saving rate based on the CHFS 2015 indicate that the competitive saving motive still holds, but only in the rural area. Specifically, in contrast to the estimates based on the CHIP 2002, in the rural, the effects on households with a son is much smaller, and the effects on families with a daughter are also significant positive. However, there are no effects on urban households according to the estimates based on CHFS 2015.

Finally, we use data from Population Census (1990, 2000, 2010), China Statistical Year-Book 1986–2016, and Comprehensive Statistical Data and Materials on 50/55/60 years of New China (CNBS) — also the same data sources as Wei and Zhang (2011a) used, and follow the same specification to reexamine the effect of sex ratio on the saving rate at province level. We adopt a more precise algorithm to compute the sex ratios of the 31 provinces for 7 years (1985, 1990, 1995, 2000, 2005, 2010, and 2015). Our results indicate that the competitive saving motive exists, but only in the rural area. The effect of sex ratio on the saving rate is insignificantly different from zero for both full sample and the urban sample. By estimation and computation, an increase in sex ratio from 1985 to 2015 can explain about 28% of the actual increase of the increase of rural saving rate.

Related literature on sex ratio After two widely cited papers Wei and Zhang (2011a) and Du and Wei (2013), Wei and Zhang (2016) has a useful survey on recent literature on the competitive saving motive and broader economic implications. As a cross country's evidence, Du and Wei (2016) report patterns of sex ratio imbalances and private-sector saving rate across countries and show that greater gender imbalance tends to correspond with the higher saving rate. The result verifies the theory presented in the Du and Wei (2013). To go beyond the cross-country evidence and examine household-level evidence is quite useful. Wei, Zhang, and Liu (2017) test the effect of local sex ratio on home ownership and home prices. The paper finds son-families in an area with more elevated sex ratio are more likely to own a home and concludes that rise in sex ratios contribute nearly half of the rise in cities' housing prices in China. Edlund, Li, Yi, and Zhang (2013) show that unbalanced sex ratios result in higher crime rates in property and indicate that rising sex ratios account for about one-seventh of overall crime increase in the period 1998–2004. Wei and Zhang (2011b) find that son-families in regions with higher sex ratios are more likely to become entrepreneurs, while daughter-families do not respond to local sex ratios. Using a natural experiment in Chinese Taiwan, Chang and Zhang (2012) and Chang and Zhang (2015) study the effect of mating competition on entrepreneurship in man's marriageable age. Using population census and elderly survey, they quantitatively indicate that young menâĂŹs entrepreneurship is inspired by such a policy change. In addition, a few recent studies focus on the evidence that son-families have incentives to earn more income. Knight, Li, and Deng (2010) demonstrate that household having a son raises household income by more than 10% using CHIP data. Ding and Zhang (2014) report that son-families receive more remittances and invest more in productive activities compared to the household with a daughter using a nationally representative rural household data in 300 villages. Li and Yi (2015) find that premarital son-families are more likely to migrate and generate more earnings than those with only a daughter and gender of the first child matters for family finance and health consequence based on China Family Panel Survey. One of the interesting research is the impact of sex ratios on happiness. Using both the China Central TV postcard survey, which includes questions on happiness and China Family Panel Survey 2000, Tan, Wei, and Zhang (2015) show that higher sex ratios consistently lower people's self-reported happiness.

Roadmap The rest of the paper is organized as following: in the Section 2.2, we replicate Wei and Zhang (2011a). Section 2.3 provides new evidence from the CHFS. We reexamine the competitive saving motive at province level in Section 2.4. Section 2.5 concludes. The Appendix contains figures, tables, and other details.

2.2 Replication of the Wei and Zhang (2011a)

CHIP 2002 For our replication, we mainly use the household-level data from Chinese Household Income Project (CHIP) of 2002, the same sample as Wei and Zhang (2011a). This CHIP data consists of three distinct samples of urban, rural and migrant household groups. The CHIP sample is selected from a larger sample drawn by the National Bureau of Statistics of the People's Republic of China and conducted by the Chinese Academy of Social Sciences in 2003. The sample cover 22 provinces, 77 urban cities, and 122 rural counties.

China population census 1990 and 2000 To reexamine the competitive saving motive, especially reconsider the county sex ratio pattern, we use the two waves of China population census in this section —in an attempt to replicate and extend Wei and Zhang (2011a) analysis. In Wei and Zhang (2011a) household-level regressions, for the cohort of ages 12–21 in the year 2002, they infer county local sex ratios from the 1990 population census of cohort 0-9 years old. Taken the data unavailability into consideration, we also include and infer the same aged cohort from the 2000 population census (the cohort was 10-19 years old in 2000). These two local sex ratio inferred are considered as two measures for the key variable of interest, and the 2000 population census is possibly closer to the true value of 2002 county local sex ratio.

It is important to point out that the Wei and Zhang (2011a)'s local sex ratio is inferred from the county report of population census 1990 which includes the specific population and detail composition of the population for all counties. That local population number is inferred from the population by the NBS. Unfortunately, county report of population census

1990 is too far for current day, since cannot be assessed to us, we use the 1990 Population Census Sample Survey Data (0.095%) to calculate the local fraction of males or females for a specific range of ages in the county and then premarital local sex ratio. There will be a little different between these two sources of sex ratio. Table A.1 report the summary statistics with the average saving rate for each household type and the comparison with the Wei and Zhang (2011a). Associated with the table 4 of Wei and Zhang (2011a), we follow the Chamon and Prasad (2010) and Wei and Zhang (2011a) to define the household saving rate as log(disposable income/living expenditure). The disposable income is the household total income net of tax. The definition has two advantages according to Wei and Zhang (2011a): one is to make the error term more likely to satisfy the normality assumption, and another is less susceptible to extreme values.

Empirical specification Following Wei and Zhang (2011a), we estimate the following empirical model

$$\texttt{saving rate}_{ij} = \beta_1 \cdot \texttt{local sex ratio}_j + \beta_2 \cdot \boldsymbol{X_i} + \beta_3 \cdot \boldsymbol{Y_j} + u_{ij}, \tag{2.1}$$

where i indicates households and j indicates counties. In this model, `saving rate` is defined as the *log* of the ratio of household disposable income to household consumption; `local sex ratio` is from the premarital cohort age form 0 to 9 in 1990 census data; $\boldsymbol{X_i}$ is a vector of household characteristics, including other determinants of the saving rate as household per-capita income, child's ages and characteristics of the household head: gender, age, age^2, year of schooling, a dummy variable denoting "poor health" if the family has severely ill member or a disabled as the health shocks to the household; $\boldsymbol{Y_j}$ refers to a vector of county characteristics. For the purpose of replication, we first control the county Gini coefficient as Wei and Zhang (2011a) did. Then we extend control variables for the counties to include county per-capita GDP, urbanization ratio, in-migration ratio, male and female unmarried ratio of age 15 and above. We also control the dummy variable defining a household without public insurance, a household with State-owned Enterprise employment, and household with a member in the reorganization firm or in a profit-loss firm or laid-off from the enterprise.

After micro-level data cleaning, we may not have the same sample used for analysis although in this replication case. Fortunately, use the similar nuclear household criteria as Wei and Zhang (2011a)[6], we get our sample very close to (Wei and Zhang, 2011a) and the summary statistics of saving rate are nearly the same. In our sample of rural area, households with a son have average and median saving rate of 39.3% and 39.4%, respectively, higher than 31.8% and 35.3% for households with a daughter. For the urban area, a household with a son and with a daughter has the similar average and median saving rate around 31% for our sample. Table A.1 also report the summary comparison for the sex ratio, in our sample inferred from the 1990 Population Census Sample Survey Data: the average sex ratio for rural counties and urban cities is 1.09 and 1.08, with a standard deviation of 0.04 and 0.04, respectively. The smallest and largest values for rural counties and urban cities are 1.01 and 1.23, along with 1.02 and 1.24, respectively. For the sex ratio sample of (Wei and Zhang, 2011a), the sex ratio ranges from 1.01 to 1.23 with a mean of 1.09 for rural counties and 1.02 to 1.24 with a mean of 1.08 for urban cities which is similar to ours. However, the standard deviation is both 0.04 for rural and urban, is much smaller than ours due to the post-adjustment for the local population by NBS.

The principal hypothesis of Wei and Zhang (2011a) focus on a particular regional variation in saving rate: holding constant household income and other family or household head level characteristics, a household with a son should save more in a region with higher local sex ratio. And the daughter may have free rider effect that not responses to the higher local sex ratio in rural counties and have to save more in a region with higher local sex ratio due to the housing prices spillover channel in urban cities.

[6]Nuclear families: a three-person household with both parents still alive, and mother's age less than 40, and no parents and other relatives living at home. The household survey cannot capture the moving out children accurately. Therefore the household is more likely to be a nuclear family by placing a limit on the mother's age. The sample size of this restricted sample is significantly smaller compared to the original sample size of CHIP survey(Wei and Zhang, 2011a), and the regression owns good statistical power.

Empirical results comparison The table A.2 present the regression results and comparison with the Wei and Zhang (2011a). Among them, the table A.2a present the rural sample regression results and comparison with the Wei and Zhang (2011a). Column 1 and 2 of table A.2a perform respectively the results comparison of the regression relating savings by a household with a son and household with a daughter on a full sample. We find the similar results that local sex ratio has a strongly positive effect on the household saving rate, which raising 1 percent of local sex ratio can increase the average rural son-household saving rate by 0.54% which is statistically significant at the 5 percent level. The magnitude of the competitive effect is smaller compared to the estimates of Wei and Zhang (2011a). However, the significance level is identical. The competitive saving motive effect is also economically large for the son-household and accounts largely in the actual increase in the average rural household saving rate. Column 2 of table A.2a report that the daughter-household does not respond to local sex ratio. The estimates of the effect are same as Wei and Zhang (2011a), and the coefficient on the local sex ratio is negative and not statistically significant. This result is also consistent with the theoretical implication in Du and Wei (2013). Same as Wei and Zhang (2011a), we remove the possible outliers through three different filters to do a sequence of additional regressions and perform and preserve the competitive saving effect in a smaller magnitude and same strong significance pattern. The same outliers removing method are also performing in the later part, and in specific, the household whose reported annual household income or consumption is taken out in the columns 3 and 4. The top and bottom 5 percent of households regarding their saving rate are taken out in the columns of 5 and 6; the top and bottom 5 percent of households in terms of their saving rate along with explicit marital status for the child are taken out in the column of 7 and 8. In all cases, we present the same patterns of results: the coefficient of the local sex ratio is around 0.5 to 0.6 and significant for son-families, but the coefficient is negative and not statistically different from zero for daughter-families.

We now turn to urban household savings, and the table A.2b present and compare them with the Wei and Zhang (2011a). In column 1 of table A.2b, we contrast the son-families result, and the coefficient on local sex ratio is positive and significant as the Wei and Zhang (2011a). In column 2, the point estimate for a household with a daughter is 0.24 and significant at the 10-percent level. We also get the magnitude of the competitive effect is smaller compared to the estimates of Wei and Zhang (2011a). However, the significance level is similar. The effect on a household with a son is larger than on daughter-families for the full sample. We also attempt to remove the possible outliers as Wei and Zhang (2011a) since the big concern for noise in the data. Through a number of same filters as rural analysis, the effect of sex ratio on son-families is consistently larger than on daughter-families, and this pattern is more robust than Wei and Zhang (2011a). The spillover effect of local housing cost on daughter-families still shows, but pressure and response are less significant than Wei and Zhang (2011a)'s estimates. In fact, in column 8, the coefficient for a household with a daughter is no longer statistically different from zero.

Do other county characteristics affect the competitive motive? The 2000 population census data not only report the county-level basic situation and the variously detailed composition of the population. And also local employment information, local urbanized population share, local migration share from other provinces or other cities, and also unmarried population share of aged 15 and above for male and female in each county. Since the CHIP wave year 2002 is very close to census survey year 2000, we can use these other characteristics at 2000 as factors in our regression to check the robustness of competitive saving motive. In addition, we also get the local GDP and local from this data source: China Statistical YearBook for Regional Economy and to check whether the economic development or economy growth can vanish the competitive saving motive.

In any case, after holding constant family income and other household characteristics, the household saving rate should respond negatively significant to local per-capita GDP. The reason is that the consumer price and household consumption structure make the living expenditure cost is higher in the economy developed county, therefore leads to a lower household saving rate. The urbanization ratio and the migration population share from other province or cities are both an indicator for local economic development, and we also expect the negative effect on household saving rate after control characteristics like household income. The share of male and female unmarried population which describe the situation and condition for the local marriage market can also be a factor for saving behavior of household with a son.

We do the same robustness check as Wei and Zhang (2011a) including that median regression, using the sample with different mother age threshold, the sample with extend household with other relatives and also using interaction term in the son and daughter full sample regression. The results are also presented in a consistent pattern: local sex ratio has significant positive effect for the son-families but not for the counterpart. To save space and focusing on our point, we don't report these tables. In table A.3, we report several additional robustness checks rather than Wei and Zhang (2011a) using the other county-level factors mentioned above. The econometric purpose is to check whether other county characteristics are omitted variables in model 2.1. We first present the results with the logarithm of per-capita GDP, and then table A.3b—A.3e report the sequence of results. In all cases, the results are still performed in the same pattern and robust. The effect of 1 percent local sex ratio increase on son-families saving rate is around 0.48% to 0.65% for the rural and around 0.16% to 0.38% for the urban. Therefore, in this subsection, the competitive saving motive still holds for son-families. The coefficient of sex ratio is around 0.2 and significant at the 10 percent level for urban daughter-families, and negative insignificant for the rural counterpart. The coefficient of the related county factors meets the expectation except for the variable "share of male and female unmarried population" with an insignificant effect, and the potential reason is the unmarried population share is not the premarital age specific.

What if the use year 2000 sex ratio rather than 1990? Unlike the county report of population census 1990 is beyond our available data, the 2000 census county report which includes post-adjustment for the local population, the detailed composition of the population for all counties by NBS is among our data availability. Another reason is that local population number of 2000 is possibly closer to the true value of 2002 county local sex ratio. Therefore, we can use sex ratio inferred from the year 2000 to do the same regression. To ensure the precision and maximize the comparability across the two waves of the population census, we match administrative division code for all counties of the year 1990 to the year 2000 due to the change of local administrative jurisdiction. The process is not that easy. We use the communique of the State Council of the People's Republic of China from the year 1990 to the year 2000, to get the history of the change of local administrative jurisdiction code, and match them together after that.

In this subsection, we use the same rural and urban CHIP 2002 sample. The only difference is the data source of sex ratio. From 1990 census data, we use the age cohort of 0-9 years old; from 2000 census data, and we use the age cohort of 10-19 years old. Therefore, the difference between population census 1990 and 2000 is mainly because of the migration for the premarital aged population between 10-19 years old in each county. A resident with more education could be more mobile such as college students with age 17 and above may stay in the city of their college after graduation. Therefore, the results may be different.

Table A.4 reports the regression results with same specification as Wei and Zhang (2011a) using sex ratio of year 2000. The top row is the result of the rural sample, and the bottom row is for urban sample. And surprisingly, the competitive saving motive effect vanish with the sex ratio of the year 2000. The coefficient of local sex ratio on the household saving rate of rural son-families is around 0.18 to 0.42 and insignificant at 10 percent level, and also no longer statistically from zero using the full sample or the sample after removing the possible outliers. The effect on urban son-families is also negligible and insignificant. The point estimates of sex ratio on a rural and urban household with a daughter are around -0.20 to 0.38 and -0.59 to -0.00 and almost insignificant, respectively. The effects vanish for CHIP 2002 both urban and rural son-families and daughter-families with 2000 China population census.

Potential explanation: migration As we mentioned above, the only difference is the data source of sex ratio, and the difference between population census 1990 and 2000 is mainly because the migration for the premarital aged population between 10-19 years old in each county. Therefore, unbalanced migrate-in and migrate-out for premarital male and female is our major potential explanation.

Table A.5 report the distribution for county sex ratio of rural counties and urban cities. Since census 2000 and census 1990 of Wei and Zhang (2011a) come from the same data source in NBS county report of the population census, we mainly compare these two distributions. The standard deviation and mean value of two groups is very close. However, since the potential premarital migration occurs within 10-years period from 1990 to 2000, the maximum value of rural counties and urban cities of the year 2000 is 1.27 and 1.23, slightly bigger than the corresponding maximum value of the year 1990. Moreover, the minimum value of census 2000 in rural counties is 0.77, which is significantly smaller than 1.01 in 1990. The pattern is also applied to urban counties with 0.92 compared to 1.02. This distribution change confirms the explanation of migration potentially.

Next, we turn to the local sex ratio comparison between the year 2000 and year 1990 using from Population Census Sample Survey Data in figure A.3. The top two sub-figures in blue show the distribution of 1990 Sample Survey and the bottom two show the distribution of county report 2000. Census Sample Survey should have larger standard deviation due to no adjustment in the local population in general. However, the minimum value of rural counties in 2000 is smaller rather than 1990 Sample Survey. The premarital female migrating-out from rural counties can explain this results. Figure A.4 present cross-section scatter plot using per-capita income calculated from rural and urban county sample in CHIP 2002 with sex ratio in 1990 and 2000, respectively. For the rural part, there is clear positive and significant relationship between per-capita income and sex ratio in 1990. However, the relationship turns to negative and significant, means that the premarital aged female tends to migrate out from poor rural counties. The plot for urban cities is, but the negative relationship is rather stronger in the year 2000. The same pattern is confirmed by the figure A.5. The higher per-capita GDP or per-capita income, the more sex ratio reducing from 2000 to 1990. To go a further step, we focus on checking competitive motive of son-families and group the son-nuclear-household sample into three parts by different criterion. The first rule is based on the per-capita income of each county. Since the CHIP data separated into the rural and urban sample, we divide the rural counties and urban cities into three parts respectively.[7] We can also group the son-families by household-income—the second rule.[8] These two pattern are presented in figure A.6 and figure A.7, respectively. Along with the unconditional effect presented, table A.6 presents the conditional causal effect on local sex ratio including the daughter-families. The top half of figure A.6 and above panel of the table A.6 present the similar correlation that for rich rural, the coefficient of county sex ratio in 1990 and 2000 on household saving rate is 1.1 and 0.9 respectively and both significant at 10 percent level. The results are robust both using the full sample or sample without outliers. However, for the poor rural counties which per-capita income is among bottom 25 percent, the competitive effect vanishes again using county sex ratio in 2000, but not for sex ratio in 1990. The bottom panel of figure A.6 presents ambiguous correlation in a similar way. Both above panel of table A.7 in rural and urban regression also reports the similar interpretation: using 1990 county sex ratio, the effects of the full rural sample and rich rural counties are positive signs and the effects are marginal significant for middle and poor rural counties. The middle and poor urban cities do present the significant competitive saving motive effect, and very rich urban cities do not. However, for local sex ratio in 2000, only rich rural and rich urban can present the significant positive effect, but other sub-sample along with the full sample give the disappearance of competitive saving motive. In specific, all daughter-families sub-sample give insignificant results except that in the poor urban cities due to the housing price spillover effect. The sub-sample analysis and comparison using the second rule are presented in figure

[7]We divide rural counties as poor rural, middle rural and rich rural counties; similarly, the urban cities can be separated into three parts that poor urban, middle urban and rich urban.

[8]We divide rural households as poor rural, middle rural and rich rural households; similarly, the urban households can be separated into three parts that poor urban, middle urban and rich urban.

A.7 and table A.7. The only sample according to middle-class household income group shows the positive effect and significant at 10-percent level with sex ratio in 2000, and the poor household and household in poor counties consistently do not respond to local sex ratio. One major explanation is following the logic from the point focusing on sex ratio comparison and the relation between sex ratio change and local economic development mentioned above. If the young female migrates out from the poor rural counties which leads to more unbalanced sex ratio in a poor county, the son-families in this county will face more unprecedented fierce competition in the marriage market. Besides, no matter where is the very poor son-families in, the household has no incentive to save. One thing, they cannot compete very well in the rich or middle cities or counties; another, the situation becomes worse in poor rural from 1990 to 2000. In an extreme case, there will be no available premarital aged female in the county; then the son-families could choose to migrate out or quit the market directly. Therefore, the migration of aged 10-19 household during the 10–years–long period could strongly affect the saving behavior and marriage motive for both son- and daughter- families. Reconsidering the sex ratio in the year 2000 is crucial for explaining the competitive saving motive and also understand the migration profile and income-competition profile in rural and urban China. Table A.8 presents one of the potential reason about the premarital female migration out from the poor rural counties. Especially in rural China, marriage migration is the major form of migration and cause of female migration. In China, daughter has less responsibility to look after elder parent compared to son, and young female has a higher probability to migrate out for working or schooling and stay in the new place for marriage. The numbers are all computed from census 1990 and Population Census Sample Survey. It implies that young women could be much more mobile than the young man because of the marriage migration in both intra-provincial or Inter-provincial way, especially for that from poor rural one. We can conclude our point that unbalanced migration of premarital male and female is the major explanation causing the competitive motive effects vanish in the full CHIP sample. The competitive saving motive could hold for the household with sufficient disposable income in the economy developed counties, but not for the poor counterpart.

2.3 Competitive saving motive: evidence from CHFS

In this section, we turn to an alternative data set — the China Finance Household Survey (CHFS) — to examine the impact of sex ratio on the household saving rate. The CHFS, conducted by Southwestern University of Finance and Economics in China, is a nationally representative longitudinal household survey data, contains sufficient information about individual and household income, expenditures, assets, and debts. It fits well with our research purpose. The survey started in 2011, and since then there are another three waves: 2013, 2015, and 2017. We mainly use the 2015 wave in this analysis. The CHFS 2015 includes 353 counties from 29 provinces (excluding Tibet Autonomous Region and Xinjiang Uighur Autonomous Region) in China and is a sample of approximately 38,000 household and 140,000 individuals.

The CHFS survey has two potential advantages. First, the survey contains much more detailed information on household finance, and also covers a much larger sample size. These advantages have important implications for the reliability of the analysis. The rich information on household finance is important for getting relative accurate income and consumption statistics; and having a large number of households means that we can control for province fixed effect to tackle with the geographical difference in economic conditions and culture, given that there emerges growing divergence across regions in China. Another advantage is the high data quality especially. The sampling design in the CHFS focuses on a large range of households whereas the CHIP sample consists of majority households from the state-owned enterprise (SOE) sector. In addition, to deal with the issue of only reporting income range in the sample (including for the high-income households), the CHFS infer the actual income from taxes reports. This approach would eliminate the effect of the top-coded, the common issue in the survey data, on the analysis.

Using the CHFS 2015 and the *local sex ratio* for the cohort of ages 10-24 in 2015 inferred from the cohort for age 5-19 years in the 2010 population census, we follow the same method and control for the same households and counties characteristics as used in Wei and Zhang (2011a) to reconsidering the evidence of the competitive saving motive. Table A.9 reports the regression results from the CHFS 2015 sample. Column 1 and 2 of the table perform the comparison results for households with a son and with a daughter on a full sample without removing possible outliers, respectively. Column 3 and 4 show the comparison results on a sub-sample with removing the bottom and top 5 percent saving rates.

The above panel of the table presents the results for the rural sample. For households with a son in the full rural sample, the estimated effect of the local sex ratio of cohort for age 5–14 on household saving rate is 0.66 and statistical significant at 5 percent level, which means that 1 percent increase in local sex ratio would lead a 0.66 percent increase on average in household saving rate. The estimate is more extensive than our replicated result using the CHIP (0.54 in Table A.2a) but is still much smaller than 1.34 in Wei and Zhang (2011a). For households with a daughter in the full rural sample, we get a more significant effect than families with a son, with raising 1 percent sex ratio would increase 0.75 percent household saving rate. By contrast, both our and Wei and Zhang (2011a)'s results in Table A.2a suggests that there is no significant effect for households with a daughter in the CHIP rural sample. Using the sex ratio of cohort for age 10–19 as an alternative measure, the effect of sex ratio on household saving rate is 0.54, although it is not statistically significant, for household with a son in the rural sample, while the effect is 0.78 for households with a daughter and it is statistically significant at 5 percent level. After removing possible saving rate outliers, the estimates in column 3 show that the effects of sex ratio for the cohort of age 5–14 on the household saving rate for a household with a son is 0.68, and it is 0.61 when using the sex ratio for the cohort of age 10–19. Both the estimated coefficients are a little bit larger than in column 1 and are statistically significant. Column 4 shows that for households with a daughter, the competitive saving motive preserve in a bit of smaller magnitude at 10 percent level.

The results for the urban sample are presented in the below part of the table. It is surprising that all estimated effects of the sex ratio on the household saving rate in column 3 and 4 become much smaller than the rural sample, and the most important is that they are no longer statistically different from zero. These results are totally different with them in Table A.2b.

2.4 Competitive saving motive: across-regions

Sex ratios and saving rates across province In order to identify the effects of the sex ratio on household saving rate at province level, we estimate the following regression equation

$$\texttt{saving rate}_{mt} = \beta_0 + \beta_1 \cdot \texttt{province sex ratio}_{mt} + \beta_2 \cdot \boldsymbol{Y}_{mt} + \lambda_t + \eta_m + u_{mt}, \qquad (2.2)$$

where m indicates provinces and t indicates years. In this model, `saving rate` is defined as the value of the aggregate household disposal income net aggregate household consumption dived by aggregate household disposal income; The income and consumption are computed from the Comprehensive Statistical Data and Materials on 50/55/60 years of New China (CNBS); $\boldsymbol{Y_m}$ refers to a vector of province characteristics, including per capita GDP (calculated from the CNBS), the young and old dependent ratio (calculated from the China Statistical YearBook 1986–2016), and the share of SOE in total labor force (calculated from the China Statistical YearBook 1986–2016); The year and province fixed effects are captured by λ_t and η_m, respectively; u_{mt} is the error term. $\texttt{province sex ratio}_{mt}$ is the variable of interest, which is calculated from the premarital cohort for age 5 to 19 (or for age 10 to 19 as a robustness check) in census data 1990, 2000 and 2010. The estimate of β_1 identifies the causal effect of the province sex ratio on the province aggregate household saving rate. The model 2.2 is estimated by the pooled OLS estimation, and the standard errors are clustered at the province level. In Table A.10, we report the main regression results. First three columns show the results with the sex ratio for age cohort 5–10 as the key regressors. In Column 1, the results presented are from the full sample. The effect of the *log* of province per capita income on the province aggregate household saving rate is 0.20 and statistically significant, which means that a 1 percent increase in the per capita income would lead to a 0.20 percent increase in the saving rate. This result is consistent with Wei and Zhang (2011a). In Table 14 of that paper, they estimate a similar coefficient. However, the estimated coefficient of the sex ratio on the province aggregate household saving rate is -0.03 and statistically insignificantly different from zero. In other words, there is no significant effect of the sex ratio on the province saving rate. This result contradicts with the finding in Wei and Zhang (2011a) where they argue that the effect of the local sex ratio on the local saving rate is significant positive with a coefficient of 0.28.

In column 2, we extend our analysis to examine whether the competitive motive exists in the urban sample. The estimated association between the local sex ratio and the local saving rate is still insignificant different from 0, although the magnitude now becomes positive at 0.05. This result reflects that in the urban sample, and the competitive saving motive may not exist. As the full sample regression, the per capita income has a significant positive impact on the local saving rate with a 1 percent increase in per capita income leading to a 0.14 percent increase in the local saving rate.

Column 3 reports the results when we restrict to the rural sample. The estimated effect of the province per capita income on the province saving rate is significantly positive. A 1 percent increase in the per capita income would lead to a higher saving rate by 0.39 percent. For the competitive saving motive, the estimated coefficient of the sex ratio is 0.24 and statistically significantly different from zero at the 5 percent level. This result indicates that the province saving rate is higher in the region with a higher sex ratio, which implies that the competitive saving rate exists in the rural.

Column 3 to 6 of the table presents the results using the sex ratio for age cohort 10–19 as the key explanatory variable. We find pretty close estimates of the coefficients for both the per capita income and the sex ratio. There is a significant positive effect of the per capita income on the saving rate, with the coefficient 0.20 for the full sample, 0.15 for the urban, and 0.39 for the rural. Although the estimated coefficients for the full and the rural sample are almost the same as using the sex ratio for age 5–19, they are insignificantly different from zero. For the rural, the estimated marginal effect of the sex ratio is 0.21 and significant at 5 percent level. Overall, the results change a little bit and support that the competitive saving motive only exists in the rural.

Discussion about the results In this section, we use data from Population Census (1990, 2000, 2010), China Statistical YearBook 1986–2016, and Comprehensive Statistical Data and Materials on 50/55/60 years of New China (CNBS) — the same data sources as Wei and Zhang (2011a) used. And we follow the same specification, but why are our estimated results different with that in Wei and Zhang (2011a)?

The main reason is probably that we use a different approach to define the sex ratio for premarital cohort and consider a shorter time span. In Wei and Zhang (2011a), they use data from 31 provinces for 27 years (1990 – 2007) in the panel data regression. Except for the sex ratio, the data contain all other variables information for each province in each year. To get the statistics for the sex ratio for each province in each year, they use the 2000 population census as a basis to infer sex ratio in other years. Specifically, they infer the sex ratio of the cohort for age 7–21 (their main focus) in 1990 from the cohort for age 17–31 in the 2000 census, and the sex ratio of the cohort for age 7–21 in 1991 from the cohort for age 18–32 in the 2000 census, and so on; similarly, the sex ratio of the cohort for age 7–21 in 2001 is inferred by the cohort for age 6–20 in the 2000 census, and the sex ratio of the cohort for age 7–21 in 2002 is inferred by the cohort for age 5–19, and so on. For the 2000 sex ratio, it is drawn directly from the 2000 census.

The major advantage of the approach in Wei and Zhang (2011a) is to enlarge the sample size. As they motioned in the paper: “Ideally, we would like to know sex ratio for a fixed age cohort in every region and every year. ... we make the following shortcut:...” Indeed, they have over 800 observations to do the panel regression. However, their approach also has a significant drawback: the algorithm assumes that there is no change in the sex ratio between 7–21 age group in the year i and [7+(2000-i)]–[21+(2000-i)] age group in 2000. It is not convincing that the cohort for age 7–21 in 1990 is the same as the cohort for age 17–31 in 2000, similarly for the cohort for age 7–21 in 2007 is the same as the cohort for age 0–14 in 2000. In fact, as discussed in the previous sections, in China's significant urbanization progress, the internal migration is one of the world's most extensive migration: from the rural to urban, or from the inland to the coast, or for the poor area to the rich area, etc. The effect of the enormous internal movement would not be ignored. Migration does affect local sex ratio. Therefore, we adopt a relatively accurate way to compute sex ratio to minimize the effect due to measurement error on sex ratio. Specifically, we consider the sample of 31 provinces only for 7 years (1985, 1990, 1995, 2000, 2005, 2010, and 2015, except for Chongqing, which has only 3 years data). The sex ratio of each province for the year 1990, 2000, and 2010 is directly drawn the corresponding census data; For the sex ratio in 1985, it is calculated from the census 1990. 1995 and 2005's sex ratio are computed by the average of before- and after- current year's sex ratio; For the sex ration in 2015, it is inferred from the census 2010. Our algorithm alleviates the concern and thus show that the competitive saving motive exists only in the rural area. By estimation and computation, an increase in sex ratio from 1985 to 2015 can explain about 28% of the actual increase of the increase of rural saving rate.

2.5 Conclusion

The high aggregate household saving rate is one of the unique features of the Chinese economy. Over the period of China rapid income growth, China's sex ratio, which is defined as the ratio of the number of boys at birth to the number of girls at birth, has also been experienced a rapid a growth. The prevailing trend has captured economists' attention on linking the saving rate to the sex ratio. Can the sex ratio explain high China's household saving rate? Starting with Wei and Zhang (2011a), they argue a explanation called "competitive saving motive" that increasingly unbalanced premarital sex ratio requires both rural and urban son-families to save increasingly more to compete in the marriage market and show that the rapid increase in premarital sex ratio during 1990-2007 can potentially explain about half of the sharp rise in China's household saving rate.

In this paper, we reexamine the impact of the sex ratio on the household saving rate. First, we use the CHIP data — the same sources of data as used by Wei and Zhang (2011a) — in an attempt to replicate and extend the estimates reported in Wei and Zhang (2011a). We use local sex ratio inferred from 2000 China population census and find the competitive saving motive only holds for the household in rich counties. The effects vanish for CHIP 2002 full sample with census data in 2000, especially for poor households and households in poor counties. We argue that unbalanced migration of premarital male and female is the major explanation for the vanishing effect.

Next, we turn to another nationally representative data set China Household Finance Survey (CHFS) 2015 wave for estimating the effect of the sex ratio on the household saving rate. The result indicates that the competitive saving motive still holds, but only in the rural area. Specifically, in the rural, the effects on households with a son is much smaller, and the effects on households with a daughter are also significant positive. However, there are no effects on urban households.

Finally, we use China population census, and provincial panel data as the same data sources as Wei and Zhang (2011a) used, and adopt a more precise algorithm to compute the sex ratios of the 31 provinces for 7 years (1985, 1990, 1995, 2000, 2005, 2010, and 2015). Our results report that the competitive saving motive exists, but only in the rural area. By estimation and computation, an increase in sex ratio from 1985 to 2015 can explain about 28% of the actual increase of the increase of rural saving rate.

3. INCOME INEQUALITY, LIQUIDITY CONSTRAINTS, AND CHINA'S HOUSEHOLD SAVING RATE

3.1 Introduction

Over the last three decades, the Chinese economy has been growing at an average annual rate of nearly 10%, and now it becomes the second largest economy in the world. One of the unique features of Chinese economy is the high and rising household saving rates: China's aggregate household saving rate has exceeded 35% in the recent decade, which is one of the highest in the world.[1] China's high household saving rate may already have real implications for the world economy. In 2005, Ben Bernanke, then a governor of the Federal Reserve Board, argued that China's large surpluses have adverse effects on richer countries' current accounts and financial markets. In fact, another unique feature of the Chinese economy over the same period should not be ignored: China's household income inequality has been among the world's worst.[2] Put them together, between 1992 and 2015, China's household saving rate has been increasing steadily from 33.98% to 37.07%, in the meanwhile, Chinese households income inequality measured by the Gini coefficient has also risen from 0.390 to 0.462.[3] Are the two simultaneously existing unique features of Chinese economy correlated? In this paper, we examine the role of income inequality interacting with liquidity constraints in explaining the high household saving rate in China.

To date, there are some compelling explanations on the "Chinese Saving Puzzle" (first refered by Modigliani and Cao, 2004) in the literature, including (1) demographic changes (Modigliani and Cao, 2004; Horioka and Wan, 2007; Curtis, Lugauer, and Mark, 2015; İmrohoroğlu, Zhao, et al., 2017; Choukhmane, Coeurdacier, and Jin, 2013; Ge, Yang, and Zhang, 2012); (2) precautionary saving motives (Meng, 2003; He, Huang, Liu, and Zhu, 2017; Chamon and Prasad, 2010; Wang and Wen, 2012); (3) gender imbalance and competitive motives

[1]See panel (1) of Figure B.1 in Appendix.
[2]See panel (2) of Figure B.1 in Appendix.
[3]See Figure B.3 in Appendix.

(Wei and Zhang, 2011a); (4) high income growth and habit formation (Horioka and Wan, 2007), co-residence effects (Rosenzweig and Zhang, 2014), financial choices (Cooper and Zhu, 2017).[4] No consensus has emerged, and the puzzle remains.

The main contributions of this paper to the literature on China's household saving rate is that we make the first endeavor to bridge income distribution with China's household saving rate and provide consistent and comforting micro-level evidence. Inspired by the literature on heterogeneous agent model in macroeconomics (Aiyagari, 1994; Achdou, Han, Lasry, Lions, and Moll, 2017), we build up a simple two-period model which links household saving rates to income inequality and liquidity constraints. Specifically, in this model, households are assumed to be different in two dimensions: (i) heterogeneity in initial wealth and flow income, with a particular case of two types, the rich and the poor; (ii) heterogeneity in time preference thereby in subjective discount factor, with a particular case of three types, the impatient, the less patient, and the patient. Also, we assume that households may face liquidity constraints.[5] Given a household's type of income and discount factor, whether the liquidity constraint is binding, consumption and saving rate are endogenously determined in the model. With this simple model, we provide several implications consistent with data facts: (1) the rich save more; (2) the proportion of constrained households for the poor is higher than that for the rich; (3) Liquidity constraints would increase household saving rate. (4) when income inequality increases, the rich save even more, in the meanwhile, the poor would also save more due to the binding liquidity constraints, and thus the aggregate household saving rate would rise.

Using three sources of large, nationally representative household survey data, the China Household Finance Survey (CHFS), the China Family Panel Studies (CFPS), and the Chinese Household Income Project (CHIP), we provide direct empirical evidence implied by the

[4]For a a comprehensive review of the facts and explanations pertaining to China's saving, see Yang, Zhang, and Zhou (2012).

[5]Stiglitz and Weiss (1981) indicate that as long as the institutional barriers (such as a lack of consumer credit, or capital market imperfections leading to credit rationing) are present, there will be liquidity constraints in the economy. Financial development in China, although has been improving over the past decades, is still underdevelopment. Thus, assuming liquidity constraints exist is reasonable.

theoretical model. First, we regress the household saving rate on current income quintile dummies to estimate the differences in the saving rate between higher income quintile and the lowest. We find a robust positive relationship between the saving rate and current income across all income classes in all three data sets. For example, for the CHFS, the estimated increments in the median household saving rate range from 30 percent in the second lowest income quintile to above 70 percent in the highest, and they are strictly increasing from the lowest income quintile to the highest. We continue to find a highly significant positive association when using subsample regressions and three years average income to correct the endogeneity problem for the current income. Estimated saving rate differences range from 35 percent to 82 percent in the CHFS for the subsample regressions, and from 5 percent to 19 percent for the average income approach. The positive relationship is even more pronounced when we exclude high-income entrepreneurs, drop younger households (below age 60), use an alternative definition of saving rate as a dependent variable, and apply per capita income to redefine income quintiles. Overall, in China, the top 20 percent permanent income households' saving rate is 19–23 percent significant higher than the bottom 20 percent households'.

We then exploit the probit regression to examine if poor households are more likely to face liquidity constraints. We use two ways to measure whether or not household i is facing a liquidity constraint, that is, the variable $\mathtt{LC}_i = 1$. One is the definition in Zeldes (1989a), which states that a household is liquidity constrained if the total value of financial assets is less than two months permanent income. Another is directly from our CHFS questionnaire, which asks respondent "Does your family have any credit cards, excluding inactivated cards?" Our estimates indicate that estimated marginal effects of the income quintiles on the probability of facing liquidity constraints range from 2 percent for the quintile 4 to 10 percent for the quintile one using measure the first definition in the CHFS. The effects are even more significant when using the second measure. In sum, the bottom 20 percent permanent income households are more likely to face a borrowing constraint, with a 12–20 percent significant higher probability.

To evaluate the effect of liquidity constraints on the saving rate, we design a difference-in-difference approach (DID) applying to the CHFS and the CFPS. Only the households that are credit constrained in 2013 (2012) from CHFS (CFPS) data are used as the whole sample. We separate them into two groups: treatment group, defined as the unconstrained households in the year 2015 (2014), and the comparison group, defined as the constrained group in the year 2015 (2014). We show that the existence of liquidity constraints would lead to a significant increase of more than 20 percent in the household saving rate.

Next, we address the research question what the general equilibrium effect of the aggregate household saving rate from a rise in the income inequality. We perform the cross-sectional regression that links the calculated county-level aggregate saving rate to the measure of income inequality for all three data sets, controlling for location fixed effects and other factors. We find, in the CHFS, that income inequality would have a significant positive impact on the aggregate household saving rate at the county level, with a 1 point on a scale of 100 measure increase in the Gini coefficient leading to an increase of 0.2 percent in the aggregate saving rate.

Finally, we provide empirical evidence that the marginal propensity to consume out of both the permanent income and transitory income would be significantly different across income classes for all three data sets. Although we do not see a diminishing MPC with income classes, there is still an essential pattern that the MPC out of both types of income for the bottom 20 percent households are much higher than that for the top 20 percent households. The estimated MPC for the top 20 percent households ranges from 200 to 400 RMB per 1000 RMB, while for the bottom 20 percent households, the range from 600 to 900 RMB per 1000 RMB.

These empirical pieces of evidence would have significant policy implications. The Chinese governmentâĂŹs policies on reducing the saving rate have not yet produced substantial results. If income inequality and liquidity constraints were the key reasons for the high aggregate household saving rate, the resulting policy would be drastically different. For example, it is appropriate for the Chinese government to design some income redistribution programs (such as EITC) to reduce income inequality or devote more resources to support the credit market development. An economic policy of tackling income inequality would lower the aggregate saving rate, thus becoming a policy of economic transition and growth.

Related literature. According to the life-cycle hypothesis (LCH), the basic idea about demographic explanation is that a decrease in the non-working population, which consists of the young and the old, would increase household savings due to the "less mouths to feed". Besides, China has a long historical tradition of children taking care of their elder parents. As a result, since the one-child policy was introduced in 1979, increased savings were not only due to the reduction in young population but also viewed as an effective substitute for children ("old-age security"). Using a ratio of working population to the number of nonworking ("minors") as a proxy to the demographic change, Modigliani and Cao (2004) find that increased China's household saving rate over the period from 1953 to 2000 can be well explained by the increased ratio of employed population to nonworking population, mainly driven by the decrease in the young dependent population. Besides, Curtis, Lugauer, and Mark (2015) conduct a quantitative overlapping generations model and also provide some evidence supporting the link between demographics and the saving rate at the aggregate level. However, applying panel data analysis and separately considering the young dependent ratio and the old dependent ratio, Horioka and Wan (2007) finds that the changes in those ratios do not go very far in explaining China's provincial household saving rate for the period 1995–2004. Using the data from the Urban Household Survey (UHS), Chamon and Prasad (2010) reach a similar conclusion: there is no significant effect of the demographic shifts in China's household saving rate. Recent work about the demographic explanations focus on bridging the micro-level mechanism with the macro-level framework and provide some micro-evidence (see, e.g., İmrohoroğlu, Zhao, et al., 2017 and Choukhmane, Coeurdacier, and Jin, 2013). One concern about the demographic explanations lies in that demographic shift is not static but dynamic. As the age population move over time, we would not see a consistently high and even rising household saving rate. In fact, since 2000, Chinese household saving rate has been rising rapidly and hit the highest point in the history.

The precautionary saving motives argue that people who are not covered by a social safety network tend to have precautionary saving motives and thus save more for unexpected events (Giles and Yoo, 2007). Although the Chinese economy has experienced rapid growth since the reform and opening up, due to lack of a safe social security and insurance network and increasing costs on education, housing, and healthcare, etc., make Chinese household tend to save more to respond the income and expenditures uncertainties in future. On the income uncertainties side, Meng (2003) examine the role of precautionary saving in Chinese urban households during the period from 1995 to 1999. She finds that not only the Chinese urban households ever experienced past income uncertainties tend to have increased propensity to consume, but for households without unemployed members, the income uncertainty has an even stronger effect on saving. Using China's reform of the state-owned enterprises (SOE) in the late 1990s as a natural experiment, He, Huang, Liu, and Zhu (2017) also show that the precautionary saving motive does exist in Chinese households. Using the CHFS, however, our preliminary results show that the saving rate of households whose heads work in government entities, public-sector organizations and state-owned enterprises are slightly and insignificantly higher (0.04% higher) than that of households whose heads work in privately-owned enterprises, collectively-owned enterprises, and foreign-funded enterprises. This result shows that China's gradually well-established labor law and law of employment contracts makes income uncertainty a less influential factor for the increasing household saving rate. On the expenditure uncertainties side, Chamon and Prasad (2010) argue that uncertainty in expenditures, particularly on education, housing, and healthcare, may generate high aggregate savings for the young and the elderly. Over the last decade or so, however, the social insurance system has been firmly established. There is almost universal health insurance coverage, and rapid retirement insurance coverage has not lowered the saving rate. Also, there is no consensus as to whether the high housing prices can explain the high household saving rate (Wang and Wen, 2012).

Another compelling explanation is the imbalanced sex ratio and competitive motive. The idea is built on the traditional culture of son preferences in Chinese households: as sex ratio increase, Chinese households tend to save more to improve son's competitiveness in the marriage market. Using household-level data, Wei and Zhang (2011a) find that saving rates for the households with sons in the high sex ratio county is significantly higher than the households with sons in the low sex ratio county in both rural and urban sample. At the provincial level, they find evidence that the sex ratio has a significant positive effect on the provincial aggregate household saving rate. They argue that during 1990–2007, the factor can account for at least 60 percent actual increase in China's household saving rate. However, we reexamine the competitive saving motives using the same data sources as they did and find the evidence may be not robust. First, although we find the similar effects of the sex ratio on household saving rates using the sex ratio from the 1990 census when using the sex ratio from 2000 census, the effects vanish. Second, even using the ratio from 1990 census, the effects exist only in the rich households sample and rich counties sample. For the poor households and poor counties, the estimates are significantly negative and statistically insignificant, respectively.

There are other explanations for Chinese household saving rate. According to Carroll and Weil (1994), the rising household savings may be due to a consequence of high-income growth and habit formation. Horioka and Wan (2007) find that the lagged saving rate has a significant positive effect on the provincial-level household saving, which is consistent with the existence of inertia or persistence. However, as argued in Modigliani and Cao (2004), during the 1950s to the mid-1970s, average Chinese household saving rate was lower than 5 percent, which implies that the Chinese cultural, ethical values of "thrifty" counts little if any. Cooper and Zhu (2017) estimate a structural life-cycle model to study household finance in China. They find that the high Chinese household raving rate is mainly driven by the labor market risk and the patient Chinese households. Other studies, such as Ge, Yang, and Zhang (2012), Rosenzweig and Zhang (2014), and Song and Yang (2012), focus on explaining another feature of Chinese household saving rate, the "U–shaped" age-saving profile started with Chamon and Prasad (2010).

Roadmap. The rest of the paper is organized as follows. Section 3.2 presents the data and stylized facts of Chinese household saving rates. In section 3.3, we introduce our theoretical model that links income inequality, liquidity constraints and saving decisions in a two-period lifecycle model. Section 3.4 describes the empirical methodology. The results and analysis are presented in Section 3.5. Section 3.6 concludes. The Appendix contains proofs, figures, tables, and other details.

3.2 Data Facts

In this section, we provide data pieces of evidence that motivate the idea that income inequality interacting with liquidity constraints matter in the explanation about China's high household saving rate, based on various household survey data sources from China. First, we plot the household saving rates by income class. The data pattern is consistent with various data sources. Second, we take a look at the household saving rate by income group and note, within each income group, the importance of facing a borrowing constraint on household saving rates. Finally, we show the relationship between the county-level aggregate household saving rate and the county Gini coefficient. These data facts motivate the theoretical model in the next section and the reduced form analysis in the following section.

3.2.1 Data sources

We use three data sets from China in the analysis. The data are drawn from the China Household Finance Survey (CHFS), the China Family Panel Studies (CFPS), and the Chinese Household Income Project (CHIP).

The CHFS is our primary source of data used in the analysis. It is a large, nationally representative and longitudinal dataset, conducted by the Survey and Research Center for

China Household Finance at Southwestern University of Finance and Economics in Chengdu, China. The survey was first launched in 2011, and another three waves were conducted in 2013, 2015, and 2017, respectively. The CHFS uses a three-stage stratified sampling method and covers 29 provinces and autonomous regions (except Tibet, Xinjiang, Hong Kong, Macao and Taiwan). It also has a low non-response rate compared to other survey data. The overall representativeness of the CHFS is excellent, and it fits our research purpose well. Besides, the survey contains detailed information about a large sample of individuals and households' demographic characteristics, assets and liabilities, insurance and social welfare, and income and expenditures. So thus the CHFS is particularly suited to our purposes. We primarily use the 2013 and 2015 waves in this study.

The CFPS, conducted by the Institute of Social Science Survey at Peking University, China, is also a nationally and representative longitudinal dataset. The survey started in 2010, and the following three waves were in 2012, 2014, and 2016. The primary purpose of the survey is to track individuals, families, and communities in contemporary China. Although the CFPS focuses on various aspects of social life, it also collects wealth information about incomes and expenditures. It fits our research purpose well. Among the current four waves, the 2012 and 2014 waves are used in the analysis.

The CHIP is nationally representative dataset conducted in 1988, 1995, 2002, 2007, and 2013. The five waves of the survey are designed to track the dynamics of income distribution of Chinese individuals and households in both urban and rural area, and thus it also contains sufficient information about incomes and expenditures. The 2013 wave is used in the study.

We apply the same criteria in all three datasets to construct our estimation sample. First, we remove outliers and households with missing data, the 2015 CHFS survey provides a sample of 21,861 urban households from 1,048 different communities in 262 counties; the 2014 CFPS survey has a total 6,603 urban sample from 1,413 distinctive communities in 358 counties; the 2013 CHIP survey data include a total 6,674 urban sample from 212 counties. Survey participation was randomized; so, again, the data are highly representative regarding the geographic location and economic development. For panel exercise in our paper, we use both three-years panel and two-years panel, and the matching households from 2013 and 2015 of CHFS data reduces the sample size to 13,120; the matching sample size is 10,677 for the 2012 and 2014 CFPS survey. We use the 2013 CHIP survey data only to perform the cross-section analysis because of the long time span for the recent CHIP survey.

3.2.2 Data evidence

Income-saving rates profile We first summarize China's uneven distribution of household saving rates across income level. Additionally, we plot Chinese households by income classes. These pieces of evidence together show the first of total four facts: not every Chinese household saves; the high-income households' savings account for a much larger fraction of total savings, with very high household saving rates. Table B.2 shows household saving rates by income classes and shares of savings for each income class, calculated from our three datasets CHFS 2015, CFPS 2014, CHIP 2013, and the National Bureau of Statistics of China (NBS). According to the CHFS 2015 data, Chinese households have an aggregate saving rate of 29.1%, which is slightly higher than the level of 28.5% from the NBS of China. It is also consistent with both the available macro data and microdata used in other studies (see Zhou, 2014 and Banerjee, Meng, Porzio, and Qian, 2014, for example). The aggregate saving rate for urban households and rural households are 37.3% and 11.6%, respectively. Also, not every household saves in the CHFS 2015 sample, however, as about 44.1% of households did not save. More important, the distribution of saving rates is extremely uneven across income classes. The top 1% of income households' total savings account for nearly 70% of total household savings, with an extraordinary high saving rate of 86.6%. The top 5% of income households have an average saving rate of 74.1%, with the share of the total savings for these households are over 99% of total savings. The saving rate for the top 10% and top 25% of income group households are 67.2% and 56.9%, respectively. As an opposite, for the bottom 50% of income households, their saving rates and the shares of savings are even negative, with -132.7% and -45.8%, respectively. The fact that household saving rate is greater for the higher income class than that for the lower income is robust in the CPFS 2014 and CHIP 2013. In the CFPS 2014, the aggregate rate goes down from 58.1% for the top 1% of income class to -45.8% for the bottom 50% income class. Moreover, the saving rate decrease to 1.7% for the bottom 50% of income class from 53.6% for the top 1% of income in the CHIP 2013. Additionally, the savings shares for the top 1% of income class own 44% of total savings in the CFPS 2014 and 12.2% of total savings in the CHIP 2013. Figure B.2 displays that Chinese households saving rate by income percentile increases as income level rises for CHFS 2015 (panel 1), CFPS 2014 (panel 2), and CHIP 2013 (panel 3).

Saving rates and credit constraints Next, we show evidence about the role of borrowing constraints on households saving rates, by comparing households saving rates for those who may be constrained with those who may be not across different income classes. In particular, we exam households whose income are above top 20%, below bottom 20%, and in the middle. For the measures of constrained household, we use the one in (Zeldes, 1989a) and the credit card usage to define whether or not a household is facing borrowing constraint. Table B.2 shows a summary for CHFS 2015, CFPS 2014, and CHIP 2013. Among 5179 urban households in CHFS 2015, there are 4330 (83.61%) households in the top 20%, 501 (9.67%) households in the bottom 20%, and 293 (5.66%) households in the middle. The first three columns in the panel (a) of Table B.2 uses the definition in the literature (Zeldes, 1989a) and the last three uses the credit card usage measure. We find that in the top 20% of income group, about 17% households are facing a borrowing constraint, and the corresponding saving rate is about 78%, which is about 40% higher than those who are not facing a borrowing constraint in the same income group. For the middle-income group, the percentage of constrained households goes up to near 32%, with the saving rate for constrained households is 13%, which is also higher than that for those unconstrained households in the group. The proportion of constrained households in the bottom 20% is even higher to 38%. The saving rate for the constrained households is still higher than that for unconstrained households, even though it is a negative number. Similarly patterns can be found in the panel (b) and (c) for CFPS 2014 and for CHIP 2014, respectively. There are more households may face a borrowing constraint for the bottom 20% of income group than that for the top 20% of income group. Moreover, within each income group, the saving rate for the constrained households tend to be higher than that for the unconstrained counterpart in each income bracket.

Aggregate household saving rate and the Gini coefficient Finally, we look at data regarding the relationship between aggregate saving rate (both the country-level and the county-level) and the Gini coefficient. Figure B.3 displays the simple time-series trend for the aggregate household saving rate and the Gini coefficient from 1994 to 2015. The household aggregate saving rate increased steadily from 34% in 1992 to 37% in 2015, in the meanwhile, the Gini coefficient rose dramatically from 0.39 in 1992 to 0.46 in 2015. Except for some periods, the Gini coefficient exhibits a similar trend to that of the household aggregate saving rate. In addition, Figure B.4 shows the simple cross-sectional patterns between the county-level aggregate household saving rate and the county-leve Gini coefficient. Panel (1) of Figure B.4 simply suggests that a county with a higher Gini coefficient may have a higher aggregate saving rate for CHFS 2015, even though the pattern seems to be not clear in the panel (2) for CFPS 2014 and (3) for CHIP 2013.

[Insert Figure B.4 here]

To summarize, in this section, as expected, we show that (1) the aggregate household saving rate is high in China; however, not all households saved, with 44% of households not saving (CHFS 2015). The distribution of household saving rate is extremely uneven. The rich tend to save more. The saving rate of the top 1% of income households is much higher than that for the bottom 50% of income group households; (2) Chinese households may face a borrowing constraint. Comparing to the top 20% of income households and the middle-income group households, there is a larger proportion of households in the bottom

20% of income group facing a liquidity constraint. Moreover, the constrained households' saving rate is higher than the unconstrained's across income classes; (3) at county-level, the aggregate saving rate may be higher for the county with higher Gini coefficient. These pieces of data patterns are robust to various sources of data sets.

3.3 A Simple Heterogeneous Agent Model of Saving

In this section, we formulate a simple two-period endowment economy with heterogeneous households in time preference, wealth and income. The parsimonious model leads to several analytic results of household saving behavior that are consistent with data evidence presented in the previous section. That is, the rich tend to save more; the poor are more likely to face a liquidity constraint; the existence of liquidity constraints leads to a higher saving rate; a higher level of income inequality may lead to a higher level of aggregate household saving rate.

3.3.1 Preference and the constraints

An individual maximizes life-time utility drawn from the consumption c_t at each period t, $t = 1, 2$

$$u = log(c_1) + \beta \cdot log(c_2) \tag{3.1}$$

where $log(c_t)$ is the per-period utility function, and $\beta \in (0, 1)$ is the subjective discount factor. The budget constraints for the household obeys

$$c_1 + w_1 = w_0 + y_1 \tag{3.2}$$

$$c_2 = (1 + R) \cdot w_1 + y_2 \tag{3.3}$$

where w_0 denotes the initial level of wealth, w_1 represents the level of wealth to carry between "now" and "the future", y_1 and y_2 are the income received today and tomorrow, respectively, and r is the net real interest rate. We assume that there is no income growth, that is,

$y_1 = y_2 = y$.

In addition to the budget constraints 3.2 and 3.3, households may also face a borrowing constraint

$$s_1 \geq -\frac{\tau \cdot y_2}{1+R} \tag{3.4}$$

where s_1 denotes the saving plan and $\tau \in [0, m]$ measures the degree of constrained.[6]

Notice that this is a two-period model in which individuals will die at the end of the second period. We do not consider the bequest saving motive in the last period (even though it could be more realistic in real life and also be important in theory). Since caring about nothing afterward, individuals will consume all the available resources in hands, and there is no savings or wealth left.

3.3.2 Heterogeneity

We assume that individuals are heterogeneous in two dimensions: (i) their initial wealth w_0 and income y_t and (ii) their subjective rate of discount factor β. First, we consider two wealth and income types, the rich (denoted by r) and the poor (denoted by p). Let w_0^k and y_t^k denote the initial wealth and income for the k class, where $k \in \{r, p\}$. And the wealth and the income for the rich and the poor satisfy

$$w_0^r = \theta_w \cdot w_0^p$$

$$y_t^r = \theta_y \cdot y_t^p$$

where θ_w and θ_y are the ratio of wealth to income for the rich and the poor, respectively.

Second, we assume that there are three types of household in terms of the subjective discount factor: impatient household with $0 < \beta_L \leq \beta_r$; less patient household with $\beta_r \leq \beta_M \leq \beta_p$; and patient household with $\beta_p \leq \beta_M < 1$.

[6]This condition ensures non-negative consumption in the second period. When $\tau = 1$, it is actually so-called "natural borrowing limit" as discussed in Aiyagari (1994).

3.3.3 Household decisions and model predictions

The individual household's optimal consumption can be solved by maximizing the lifetime utility function 3.1, subject to budget constraints 3.2 and 3.3, and the borrowing constraint 3.4, given the exogenous wealth and income, the discount factor, and the interest rate. Since there are three levels of subjective discount factor and two levels of initial wealth and income, there are total six types household in this model economy: (1) rich–impatient household, (2) rich–less patient household, (3) rich–patient household, (4) poor–impatient household, (5) poor–less patient household, and (6) poor–patient household.

For each type of household, the interior solution (that is, the borrowing constraint is not binding) for optimal consumption satisfies the intertemporal Euler equation, and the assumption of log utility function implies that the current optimal consumption is a linear function of the present value of lifetime resources, with a fixed proportion

$$c_1^k = \frac{1}{1+\beta^j}(w_0^k + y^k + \tau \cdot \widetilde{R} \cdot y^k). \tag{3.5}$$

It follows from Equation 3.5 that household's current optimal saving is

$$s^k = \left(1 - \frac{1+\tau \cdot \widetilde{R}}{1+\beta^j}\right) y^k - \frac{1}{1+\beta^j} w_0^k. \tag{3.6}$$

If the borrowing constraint is binding, then the kinky solution for the current optimal consumption is

$$c_1^k = (1 + \tau \cdot \widetilde{R}) \cdot y^k. \tag{3.7}$$

Whether or not the household is facing the borrowing constraint depends on wealth/income group and patient degree. To characterize the equilibrium, we introduce the following assumption

Assumption 1. *(a) Assume that the initial wealth gap between the rich and the poor is smaller than the permanent income gap between them, that is,* $\theta_w/\theta_y < 1$.

(b) The cutoff values of β *are given by following equations*

$$\beta_r = \frac{\rho_r + 1 - \tau \cdot \widetilde{R}}{1 + \tau \cdot \widetilde{R}}$$
$$\beta_p = \frac{\rho_p + 1 - \tau \cdot \widetilde{R}}{1 + \tau \cdot \widetilde{R}}$$

where $\rho_k = \frac{w_0^k}{y_1^k}$, $k \in \{r, p\}$ *and* $\widetilde{R} = \frac{1}{1+R}$.

This assumption implies that the initial wealth-permanent income ratio for the rich is smaller than that for the poor, that is, $\rho_r < \rho_p$. Figure B.5 displays the data evidence that motivates us. Optimal consumptions and savings for different types of household can be summarized in the following proposition.

Proposition 1. *Under Assumption 1, (1) for the rich-impatient, the poor-impatient, and the poor-less patient household, the borrowing constraint is binding and thus the current optimal consumption*

$$c_t^{r,L} = (1 + \tau \cdot \widetilde{R}) \cdot y^r, \quad \textit{for the rich-impatient household},$$

and

$$c_t^{p,j} = (1 + \tau \cdot \widetilde{R}) \cdot y^r, \quad \textit{for the poor-j household}, j \in \{L, M\};$$

(2) for the rich-less patient, the rich-patient, and the poor-patient household, the borrowing constraint is not binding and thus the current optimal consumption

$$c_t^{r,j} = \frac{1}{1+\beta^j}(w_0^r + y^r + \tau \cdot \widetilde{R} \cdot y^r), \quad \textit{for the rich-j household}, j \in \{M, H\},$$

and

$$c_t^{p,H} = \frac{1}{1+\beta^H}(w_0^p + y^p + \tau \cdot \widetilde{R} \cdot y^P), \quad \textit{for the poor-patient household.}$$

These analytic solutions are useful to convey simple predictions that are consistent data evidence presented in section 3.2. Those model predictions are summarized as follows:

1. The rich tends to save more.
2. Among the poor household, there is a larger fraction of households facing the borrowing constraint than that among the rich households.
3. The existence of liquidity constraints lead to a higher aggregate household saving rate.
4. Increasing the current income inequality would make the aggregate household saving rate even higher.

3.4 Empirical Strategies

In light of the theoretical model in Section 3.3, we construct and estimate several empirical models to study: (i) do the rich save more? (ii) are the poor more likely to face the liquidity constraints, and do the borrowing constraint leads to a higher saving rate? (iii) whether or not, at the aggregate level, income inequality will have a positive effect on the household saving rate? (iv) does the marginal propensity to consume (MPC) decreases with income level?

3.4.1 Income and saving rate

Econometric specification. Following Dynan, Skinner, and Zeldes (2004), we consider the following empirical specification

$$\texttt{saving rate}_i = \alpha + \boldsymbol{\beta} \cdot \boldsymbol{D}_{\text{incG}_i} + \boldsymbol{\gamma} \cdot \boldsymbol{X}_i + \epsilon_i. \tag{3.8}$$

In this model, the dependent variable is the household saving rate, which is defined as the ratio of household disposable income minus household consumption to household disposable income.

The explanatory variable of interest $\boldsymbol{D}_{\text{incG}_i}$ is a vector of dummy variables for income quintile that take a value of one if the household's income belongs to specific income quintile and zero if the ith household's income is not in this quintile. These dummy variables capture the different types of income class as discussed in the simple theoretical model in Section 3.3. The regression model in equation (3.8) also includes some control variables that capture household characteristics and household head's characteristics. These independent variables are used to control for other saving motives in the existing literature. The regression errors are denoted by ϵ.

The regression model (3.8) is estimated by running the mean and the median cross-sectional regression. In each case, we include dummies for all income quintiles except for the first one. The critical parameters of interest are the coefficients of the income quintiles. Each estimated β for a given income quintile captures, all else equal, the average excess saving rate for households in that quintile relative to households in the last income quintile.

The measurement. The variables used in our analysis include household consumption, household income, household demographic variables (household size, young dependent ratio, old dependent ratio), precautionary-type variables (employed status, employed type, hukou, health status, health insurance, pension, housing), competitive-type variables (number of boys and girls, age of children), and a set of household head characteristics (age, gender, married status, ccp member, years of schooling). The detailed definition of each variable is shown in panel (a) of Table B.3. Panel (b) - (d) of Table B.3 shows the summary statistics of these variables in the full sample. **Endogeneity.** One problem in the regression (3.8) is the correlation between `current income` and the error term. According to Friedman (1957), one's consumption at a point in time does not only depend on the current income, but also on the permanent income, the expected long-term average income. To solve the endogeneity issue of current income, we adopt two approaches. First, since the previous related literature has typically found that the association between current income and permanent income will become close to one between one's mid-thirties and forties (see Haider and Solon, 2006, Böhlmark and Lindquist, 2006, and Grawe, 2006), we do the regression on a subsample which is restricted to include those households whose head's age is between 30 and 45. Obviously, this approach will suffer a dramatic decrease in sample size. Second, we deal with the endogeneity issue by constructing the measure of permanent income following the most applicable approach in Fuchs-Schündeln and Schündeln (2005) and Bhalla (1980). Specifically, we use an average of the current income and the recent past incomes as a proxy for the measure of permanent income, and then re-group households by using quintile based on the measure of permanent income.

Robustness check. To examine to what extent the analysis results are robust, we conduct several robustness checks. First, using the definition of non-entrepreneurs in Gentry and Hubbard (2000) (the value of business income for a household is less than $5,000), we restrict our samples to non-entrepreneurs so that we can test if the saving behavior of entrepreneurs drives the main results.[7] Second, we consider if the relationship between income and saving rate is consistent for the older ages population, by restricting our sample to households with household head's age above sixty. Third, rather than define income quintile by household income, we use per capita household income to regroup households as income quintile dummies to check the robustness of the estimates. Finally, we investigate whether or not the main results are robust by using an alternative definition of household saving rate introduced in Chamon and Prasad (2010) and Wei and Zhang (2011a).

3.4.2 Liquidity constraint and saving rate

In Section 3.2, we present data evidence that across three sources datasets, there are more households may face a borrowing constraint for the bottom 20% of income group than

[7] In the literature, Quadrini (1999, 2000), Gentry and Hubbard (2000), and Hurst and Lusardi (2004) have emphasized that the high-income entrepreneurs plays an important role in wealth accumulation and thus in savings.

that for the top 20% of income group, and within each income group, the saving rate for the constrained households tend to be higher than that for the unconstrained counterpart in each income bracket. In this subsection, we formally estimate the probability gap of facing liquidity constraints between income quintiles and the effect of the liquidity constraints on the household saving rates.

First, we run a probit regression to estimate the difference in probability of facing liquidity constraints between income quintiles

$$\Pr(\mathtt{LC}_i = 1) = \Phi(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{D}_{\mathrm{incG}_i} + \boldsymbol{\gamma} \cdot \boldsymbol{X}_i), \tag{3.9}$$

where $\mathtt{LC}_i$ is a dummy variable, which takes a value of one if household i face a borrowing constraint and zero if the household does not face a borrowing constraint. We use two ways to measure whether or not household i is facing a liquidity constraint, that is, the variable $\mathtt{LC}_i = 1$. One is the definition in Zeldes (1989a), which states that a household is liquidity constrained if the total value of financial assets is less than two months permanent income. Another is directly from our CHFS questionnaire, which asks respondent "Does your family have any credit cards, excluding inactivated cards?" $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The explanatory variable $\boldsymbol{D}_{\mathrm{incG}_i}$ and control variables $\mathbf{X}_i$ are the same as equation (3.8). When do the regression, the last income quintile is omitted so that the coefficients of interest $\boldsymbol{\beta}$ are used to measure the difference in probability of facing a borrowing constraint relative to the last income quintile. We also use three-year moving average income as a proxy to permanent income to solve for the endogeneity problem in the probit regression.

Next, we use a Difference-in-Difference (DID) method to estimate the effect of the liquidity constraints on the household saving rates. Specifically, we estimate the following model

$$\texttt{saving rate}_{it} = \alpha + \beta \cdot \texttt{credit}_{it} \cdot D_{it} + \gamma \cdot D_{it} + \rho \cdot \texttt{credit}_{it} \qquad (3.10)$$
$$+ \delta \cdot \texttt{household income}_{it} + \boldsymbol{\mu} \cdot \boldsymbol{X}_{it} + \epsilon_{it}, \quad t = 1, 2,$$

where $\texttt{credit}_{it}$ is a dummy variable which takes a value of one if household i is credit constrained in the first period but unconstrained in the second period and zero otherwise. D_{it} is a time dummy variable that takes a value of one if $t = 2$ and zero if $t = 1$. `household income` is household's total disposal income, and $\boldsymbol{X}$ includes other household and household head's characteristics that are the same as previous regression equations. ϵ refers to regression errors. The key coefficient of interest is β, which measures the average "treatment" effect of credit unconstrained on the household saving rate.

The "common trend" assumption is required to be held for identifying β in DID approach. That is, in the presence of financial constrained, households without financial constrained would have experienced changes in saving rate similar to those with financial constrained. We address the validity of this assumption by using the t-test of all the controlled characteristics of the treatment and comparison group since we only have one-year panel sample before the treated year. After the control variables t-test for these two groups, all the indicators on the leading year are not statistically different between treated and comparison group; therefore, it provides support for the validity of the identifying assumption.

3.4.3 Income inequality and saving rate

In order to identify the effects of income inequality on household saving rate at aggregate level, we estimate the following regression model

$$\texttt{county saving rate}_i = \alpha + \beta \cdot \texttt{Gini}_i + \boldsymbol{\gamma} \cdot \boldsymbol{X}_i + \epsilon_i, \qquad (3.11)$$

where the dependent variable is `county saving rate`, which is defined as the ratio of the sum of total household savings in the same county to the sum of total household disposable income in the same county. `Gini` is the Gini coefficient at county-level, and $\boldsymbol{X}$ contains other county characteristics. ϵ refers to regression errors. The key coefficient of interest is β, which measures the average effect of income inequality on the aggregate saving rate at county-level.

We focus on the Gini coefficient as the primary measure for income inequality, and use other measures, including the income ratio of the top 20% of households to the bottom 20% of households as a robustness check. Control variables in $\boldsymbol{X}$ include other county demographics such as *log* of county per capita income, county level young dependent ratio, and county level old dependent ratio, and etc.

3.4.4 Marginal propensity to consume (MPC)

The old idea that the marginal propensity to consume out of current income is diminishing, that is, consumption function is concave in current income, can be dating back to the discussion in Keynes (2016), which writes "...not only is the marginal propensity to consume is weaker in a wealthy community, but, ..." in part II of Chapter 3, and "But with the growth of wealth and the diminishing marginal propensity to consume, ..." in part V of Chapter 23. On the theoretical side, a formal analytical explanation for the intuition does not appear until Carroll and Kimball (1996).[8] On the empirical side, there is little recent literature that provides the micro empirical evidence that the marginal propensity to consume for the rich is lower than that for the poor. Lusardi (1996) estimates the changes in household consumption response to the changes in transitory income using two panel data sets and provides evidence of the concavity of the consumption function, and Souleles (1999) examines the response of household consumption to income tax funds and finds that the

[8] Carroll and Kimball (1996) show that a sufficient condition for the concavity of consumption function, in most of the cases, is that introducing income uncertainty into the utility maximization problem. Before them, Zeldes (1989b) uses numerical methods to show adding labor income uncertainty can make consumption function concave, and Kimball (1990) explains the increase in the slope of the consumption function.

response is smaller for rich households.[9]

In this subesection, rather than estimating the changes in consumption response to the changes in income, we estimate directly the marginal propensity to consume out of current income following Paxson (1992). Specifically, we estimate the following consumption function

$$\texttt{consumption}_i = \alpha + \beta \cdot \texttt{transitory income}_i + \gamma \cdot \texttt{permanent income}_i + \boldsymbol{\delta} \cdot \boldsymbol{X}_i + \epsilon_i, \tag{3.12}$$

where `consumption` is household total consumption, `transitory income` is the measure of household temporary income in current period, which equals to the difference between `current income` and `permanent income`, `permanent income` refers to the household's expected long-term average income, and $\boldsymbol{X}$ summarizes household characteristics and household head's characteristics. The key coefficient of interest is β, which measures, by definition, the marginal propensity to consume out of transitory income.

For the measure permanent income, as in the previous subsection, we use the moving average of three years income as a proxy for the permanent income. The equation (3.12) is estimated by running five separate cross-sectional regressions, with each one focusing on an income quintile sample.

3.5 Empirical Results

In this section, we present our main results to answer four empirical questions: (1) whether rich households tend to save more? (2) whether poor households are more likely to face liquidity constraints, and the liquidity constraint makes the household saving rate higher? (3) whether income inequality, at the county-level, has a positive effect on aggregate household saving rate? (4) whether the marginal propensity to consume for poor households is higher?

[9] There is an older literature focusing on the test of the hypothesis of permanent income, which also finds the concavity evidence, e.g.

3.5.1 Income and saving rate

The empirical model in equation (3.8) is estimated using the CHFS 2015, CFPS 2014, and CHIP 2013 data. The coefficient of interest is the parameter of the income quintiles $\boldsymbol{\beta}$, which measures the additional saving rate for each income quintile relative to the first one when controlling for the effects of all other factors that may also affect the household saving rate. The estimates in Table B.4 are estimated by median regression, and the standard errors for the coefficients are achieved in parentheses by bootstrapping based on 500 replications.

Saving rate and current income Panel (a) of Table B.4 presents our estimation results with the first two columns showing the estimated effects for CHFS 2015 Urban, the next two columns for CFPS 2014 Urban, and the last two columns for CHIP 2013 Urban. Odd columns show estimates without controlling for the variable `employ_typ`, while the estimates in even columns controlling for it.

Column 1 and column 2 of the panel (a) suggest that the household saving rate increases dramatically with measured current income in the CHFS. The increments in the median household saving rate range from 30 percent in the second lowest income quintile to above 70 percent in the highest, and they are strictly increasing from the lowest income quintile to the highest. All the differences in these columns are statistically significant at 1 percent significant level. We also report the estimates for other factors that may affect the household saving rate. Among demographic-type factors, one increase in household size (`hh_size`) would decrease household saving rate by 2 percent, and a 1 percent increase in the old-dependent ratio (`ODratio`) would increase 0.07–0.10 percent saving rate. Both estimates are significant. These results are consistent with the existing demographic explanations about Chinese household saving rate. In precautionary-type variables, a household with urban hukou would decrease saving rate by 2 percent in column 1, although it is not significant when controlling for employ type `employ_typ` in column 2. The private burden of possible expenditures on health and housing would have a positive effect on the household saving rate in column 1, which is also consistent with the explanation focusing on precautionary motives.

Column 3 and column 4 shows results from similar regressions using the CFPS data. The estimates of the coefficient between the household saving rate and current income are smaller than in the CFPS. Nevertheless, we still see the estimated the differences in median saving rate rising significantly from 25 percent for households in the bottom quintile to over 50 percent of households in the top quintile. The qualitative effects of other factors such as household size, hukou, health, and housing on the saving rate are robust, although the magnitude is different.

The remaining columns of the table show the relationship between the household saving rate and current income in the CHIP data. As in the CHFS and the CFPS, the change in the saving rate strictly increases as income quintile moves up. For the second lowest-income households, the estimated median saving rate is 7–8 percent higher than the lowest-income, and for the highest income quintile, it is 25 percent higher than the lowest. Although the estimates are much lower than the comparable numbers from the CHFS and the CHIP, the result is not surprising: the variation in household income is much smaller in that the CHIP covers a more substantial proportion of households working in the state-own enterprise (SOE).

Endogeneity: saving rate and permanent income We now adopt two approaches described in subsection 3.4.1 to investigate the relationship between the household saving rate and permanent income. We first do the regression on a subsample which is restricted to include those households whose head's age is between 30 and 45. Obviously, this approach will suffer a dramatic decrease in sample size. Then, we use an average of the current income and the recent past incomes as a proxy for the measure of permanent income and re-group households by using quintile based on the measure of permanent income. The results are presented in panel (b) Table B.4, with the first two columns showing the estimated results for CHFS, the next two columns for CFPS, and the last two columns for CHIP.

The odd columns of the table show that when subsample regression is used to consider the endogeneity issue. The estimated change in the saving rate increases consistently with income level for all three data sets. Indeed, the difference in the saving rate is significantly positive for every quintile at 1 percent significant level, range from 35 percent to 82 percent in the CHFS, 26 percent to 62 percent in the CFPS, and 8 percent to 23 percent in the CHIP. The estimated gradients of the coefficients are similar to (and in some cases slightly larger than) to those in panel (a), the relationship between the saving rate and the measured current income.

Our next approach is to use a three-year moving average current income as a measure of the permanent income. For the CHFS, we do have three years income data surveyed in 2010, 2012, and 2014. For the CFPS, we also have income data surveyed in 2009, 2011, and 2013. Although the CHIP data cannot keep track of households over time because it does not have a panel dimension, it contains information about previous incomes, and thus we can still calculate a three-year average income. The results are reported in the even columns of the table. This procedure also yields a strong relationship between the saving rate and the permanent income. For the CHFS, the estimated differences in the saving rate range from 5 percent for the second lowest-income quintile to 19 percent for the highest-income quintile. Except for the estimated differences for the second lowest-income quintile, the CFPS also shows a highly significant correlation, with the range the range from 14 percent for the third income quintile to 17 percent for the top income quintile. For the CHIP, the estimated changes in the saving rate are all significant, with the range from 8 percent for the bottom 20 percent of households to 23 percent for the top 20 percent. In all cases, we again see the saving rate strictly increases with the predicted permanent income. What's more, the magnitudes of the saving rate with respect to income are quite close across the three data sets, but the estimates are much smaller than those in panel (a) and the odd columns. This result suggests that much of the effects of transitory income is eliminated when simply using a three-year average as a proxy to permanent income. **Robustness checks** We now turn to show several tests to check the robustness of the main results. We first extend the analysis to explore the extent to which dropping all entrepreneurs with business income is greater than $5,000 (Gentry and Hubbard, 2000) affects the results. Next, we present the results where we restrict the sample to households at older ages (above 60) from the analysis. Finally, we investigate whether the effects are robust by using an alternative definition of household saving rate introduced in Chamon and Prasad (2010) and Wei and Zhang (2011a) and by constructing quintile dummies using per capita income, respectively.

In panel (c) of Table B.4, we present the estimates based only on non-entrepreneurs and older households. Odd columns in the table continue to show a highly significant positive correlation between the median household saving rate and the income quintile, with the range of the differences in saving rate from 44 percent for the second lowest-income quintile to 91 percent to the highest in the CHFS, from 23 percent to 60 percent in the CFPS, and from 9 percent to 24 percent in the CHIP. When considering only the older households, the estimated differences in the even columns are still significant positive and strictly increasing, with the gradients much higher for the CHFS and the CFPS and similar to the previous results for the CHIP. The estimated differences rise from above 70 percent for the second income quintile to above 140 percent for the fifth in both CHFS and CFPS, whereas from 8 percent to 27 percent in the CHIP. These even higher estimates suggest no evidence that at the older age, high-income households dissave at a faster rate than low-income households.

Panel (d) of the table shows the results using an alternative definition of saving rate (in the odd columns) and alternative income quintile (in the even columns). Again, the results show that there is a strong positive association between the saving rate and income in all three data sets, and the estimates strictly income with income level. The CHFS has the largest estimated coefficients in both odd columns and even columns, with the range from 33 percent for the second lowest-income quintile to 121 percent for the highest-income, while the chip's estimated coefficients are smallest in both odd columns and even columns, with the range from 8 percent for the second quintile to 25 percent for the fifth quintile.

The validity of the theoretical model Finally, we present the empirical counterpart of the assumption in Section 3.3 and examine the validity of the theoretical model. Since we do have data on previous period wealth, there are only results for the CHFS and CFPS in panel (e) of Table B.4, with first three columns are results for the CHFS, and the last three are for the CFPS.

The first column is estimated from the similar regression to equation 3.8 with restricting the sample to households having data on previous period wealth in the CHFS. The estimated differences in the saving rate are all significant positive and strictly increasing. We report the estimated differences in the ratio of previous wealth to current income for income quintiles in column 2. The results show that the changes in the ratio strictly decrease as income quintile moves up. The estimated ratio is six times lower than the bottom 20 percent for the second lowest-income quintile, ten times lower for the third income quintile, 12 times lower for the second highest, and 15 times lower for the highest. This suggests that the assumption is realistic and reasonable. After controlling for the wealth-income ratio, in column 3 of the table, the effect of the ratio on saving rate is significantly negative. Besides, we see again highly significant and strictly increasing estimated differences in saving rate for income quintile. The smaller coefficients for the income quintiles along with a significant negative coefficient for the ratio implies that the previous wealth-current income ratio may be a channel in explaining why the rich tend to save more. As in the CHFS, similar patterns can be found in the CFPS, and they are shown in column 4–6.

We summarize the results presented so far: looking at all three data sets, although the estimates in magnitude of the increments in median household saving rate for different income quintiles relative to the bottom 20 percent differ, the pattern is generally the same —— as income quintile moves up, the difference in the saving rate between the higher income quintile and the lowest income quintile is strictly increasing, which implies that the rich do save more.

3.5.2 Liquidity constraint and saving rate

After carefully examining the relationship between the income distribution and household saving rate, in this subsection, we formally illustrate that the bottom 20% households are significantly more likely to face a borrowing constraint than the richer household and that borrowing constraint will lead to about 20% of household saving rate increases. We use two ways to measure whether household i is facing a liquidity constraint, that is, the variable $\texttt{LC}_i = 1$. For one measurement, we can use the financial liquidity constraint measure of Zeldes (1989a) in our three sources datasets. Zeldes (1989a) is the first paper using this financial constraint measure, "a household is liquidity constrained if the total value of financial assets is less than two months permanent income," and applying it into the PSID data. The paper finds that the consumption growth responds very strongly to lagged disposable income for the household with low wealth and also find that the similar estimated responses were sometimes statistically insignificant and smaller. He interprets this results as evidence that in favor of liquidity constraints and between 30 to 66 percent of households in PSID sample are liquidity constrained by this measurement and different definitions of "low wealth". For another measurement, we use it directly from our CHFS questionnaire, which asks respondent "E2002: Does your family have any credit cards, excluding inactivated cards?". For the households who reply the question "Yes" as the answer will not face the borrowing liquidity constraint, and the counterpart will face the borrowing liquidity constraint, that is $\texttt{LC}_i = 1$, due to the household do not own an activated credit card. Since only the China Household Finance Survey ask the respondent question about the credit card information, we only use CHFS 2015 data to examine the relationship between the borrowing constraint and income distribution.

Liquidity constraints and current income In this subsection, our dependent variable is the probability of $\texttt{LC}_i = 1$, whether household i is facing a liquidity constraint, and the explanatory variable of interest is also the income quintile dummy variables vector $\boldsymbol{D}_{\text{incG}_i}$. We omit the highest income quintile for the reference in each regression and directly report the marginal effects of each explanatory variable in the results table for the explanation convenience. Both the results from using the household current and permanent income, we illustrate similar evidence suggest that financial constraint and borrowing constraints are important for the poor household in China which is consistent with the evidence of U.S. (Zeldes, 1989a).

First column of panel (a) and the odd columns of panel (b) of Table B.5 present our estimation results showing the estimated effects of the current income distribution on the first measurement of financial liquidity constraint. Column 1 of the panel (a) suggest that the probability of the household facing the binding financial constraint increases dramatically with measured current income in the CHFS. The increments in the probability of facing financial liquidity constraint range from 2 percent in the second highest income quintile, 5 percent in the third highest income quintile, 9 percent in the second lowest income quintile to 10 percent in the lowest income quintile all compared to the highest quintile of current income. Except for the coefficient of second highest income quintile, all the difference in this column is statically significant at 1 percent significant level. The poor are significantly more likely to face the liquidity constraint, and this relationship pattern is the same for CFPS and CHIP data. The odd columns of panel (b) of Table B.5 shows that the increments in probability of facing financial liquidity constraint range from 5 to 6 percent in the second highest income quintile to 23 to 13 percent in the lowest income quintile all compared to the highest quintile of current income for the CFPS and CHIP data respectively. Almost all the difference in this column are statically significant at 1 percent significant level. Therefore, the estimates pattern in magnitude of the increments in probability of facing the liquidity constraint using the current income quintile is generally the same: as income quintile moves down, the difference in the probability of facing the liquidity constraint between the higher income quintile and lowest income quintile is strictly increasing, which implies that the poor are more likely to face the financial liquidity constraint.

For the current income distribution on the second measurement of borrowing liquidity constraint, column 3 of the panel (a) present our estimation results showing the estimated effects. Column 3 of the panel (a) suggest that the probability of the household facing the binding borrowing constraint increases dramatically with measured current income in the CHFS. The increments in the probability of facing borrowing liquidity constraint range from 13 percent in the second highest income quintile, 18 percent in the third highest income quintile, 23 percent in the second lowest income quintile to 27 percent in the lowest income quintile all compared to the highest quintile of current income. All the difference in this column is also statically significant at 1 percent significant level. The poor are significantly more likely to face the binding borrowing constraint. The magnitude of the effects on borrowing constraint is much larger compared to the first financial measure using the CHFS data. The reason is that the credit card application in China has multiple criteria such as the stable income with good credit record or history, enrolled in the social security system and working job type. A household with "self-employed", "owning small private business" or "farmers" are all very difficult to apply for the credit card in China financial institution. So the variance of the second measurement variable is larger than the first one, the difference of facing the borrowing constraint from the financial institution is more important and severe for the poor household. Therefore, the difference in the probability of facing the borrowing constraint, measured as "without an activated credit card," is strictly increasing as income quintile moves down, which also implies that the poor are more likely to face the borrowing liquidity constraint. **Liquidity constraints and permanent income** These above estimates may be potentially problematic if there is a third factor that varies across the population and is driving the difference between the probability of facing the financial liquidity constraint and current income or if there exists reverse causality. For example, although the current income level of a household will lead to the liquidity constraint degree, facing the binding borrowing or financial liquidity constraint will also lead to the short-term property and total income reduction. Therefore, we also use the household permanent income to correct this endogeneity, and the permanent income is defined same as the above in three datasets.

Column 2 of the panel (a) suggest that the probability of the household facing the binding financial constraint increases dramatically with permanent income defined same as the above in the CHFS. The increments of magnitude and significance in probability of facing financial constraint with the permanent income quintile is similar to the current income, range from 1 percent in the second highest income quintile, 4 percent in the third highest income quintile, 7 percent in the second lowest income quintile to 12 percent in the lowest income quintile all compared to the highest quintile of permanent income. The magnitude of this pattern is slightly larger for the second measurement of borrowing constraint column 4 of panel (a): from 7 percent in the second highest income quintile, 12 percent in the third highest income quintile, 16 percent in the second lowest income quintile to 17 percent in the lowest income quintile all compared to the highest quintile of permanent income. The magnitude of the effects on borrowing constraint is smaller compared to the current income quintile analysis using the CHFS data. All the differences in this column are statically significant at 1 percent significant level. Similarly, the pattern also holds or CFPS and CHIP data. The even columns of the panel (b) of Table B.5 also shows the increments effects range from 5 to 6 percent in the second highest income quintile to 20 to 15 percent in the lowest income quintile all compared to the highest quintile of permanent income for the CFPS and CHIP data respectively. Almost all the difference in this column are statically significant at 1 percent significant level. Therefore, the difference in the probability of facing the financial or borrowing constraint is strictly increasing as permanent income quintile moves down. Since the permanent income for the household can be treated as exogenous in the current financial or borrowing constraint situation, these results confirm our theoretical implication along with the data facts that the poor household are more likely to face the financial and borrowing liquidity constraint than the counterpart in China.

We also report the estimates for other factors that may affect the probability to face the financial or borrowing liquidity constraint both for the current income and permanent income specification. In demographic-type variables, 1 percent increase of the young-dependent ratio (`YDratio`) will decrease the probability of facing the liquidity constraint by 0.22 to 0.24 percent. Among precautionary-type factors, the registered residence status (`hukou`) change form the rural *hukou* to the non-rural *hukou* will decrease the probability of facing the liquidity constraint by 7 to 8 percent, which is consistent with the explanation that *hukou* restrictions system can depress private consumption demand of the migrant or the non-rural *hukou* household in the urban city. In addition, the household which has a poor health person or not being enrolled in public or private health insurance or pension insurance system will increase the probability of facing the liquidity constraint by 6 to 7 percent, 3 to 5 percent and 12 to 13 percent, respectively, taken the household social security into consideration. The estimated age-profile of the probability of facing the liquidity constraint is like hump-shaped, as the estimated coefficient of age (`age`) and age square ($\texttt{age}^2$) is significantly positive and significantly negative, respectively. A married household will be 4 to 5 percent significantly less likely to face the liquidity constraint due to the complete family structure and the China Communist Party membership of the household head will be 1 to 3 percent significantly less likely to face the liquidity constraint. Moreover, the household head with one more year of schooling will decrease the probability of facing the liquidity constraint about 2 percent. All the above-estimated coefficient of these other factors are statistically significant from zero and consistent with consistent with the explanation of previous literature.

Liquidity constraints and household saving rate Similar to the Zeldes (1989a) as in the U.S., we next to issue that the liquidity constraint is important to understand the poor household consumption and saving behavior. Therefore, in this subsection, we identify the effect of liquidity constraint on the household saving rate in China. In fact, in the previous specifications on identifying the income distribution and saving rate, we control our liquidity constraint measure and find the significant negative effects if the household has the financial credit or an activated credit card; however, these estimates may be potentially problematic if there exists reverse causality. For example, the current income and saving level of a household will definitely affect the next period income and also affect the household financial and credit market behavior. Also, a third unobservable factor such as the risk and financial attitude that drives the difference between the household liquidity constraint and consumption behavior. Therefore, we use a Difference-in-Difference (DID) method to estimate the effect of the liquidity constraints on the household saving rates.

We only use the household that is credit constrained in 2013 as the whole sample from CHFS data and separates them into two groups: treatment group is the unconstrained household in the year 2015 and comparison group is the still constrained group in the year 2015. We use the Difference-in-Difference (DID) specification, and the variable of interest is the interaction term of treatment dummy and year dummy variable, which measures the average "treatment" effect of credit unconstrained on the household saving rate.

We use the t-test of all the controlled characteristics of the treatment and comparison group in the leading year to address the validity of the "common trend" assumption. AS the results, we provide support for the validity of this identifying assumption. Thus, according to the Difference-in-Difference (DID) specification, the key coefficient of interest is that before the interaction term of treatment dummy and year dummy variable. As the column 1 and 2 of table B.6 shows, the estimated effects of credit unconstrained on the household saving rate is negative significant by using the credit card measure. The difference of odd and even column is that the income control variable, for the odd ones we simply control for the household disposable income, and control for the logarithm of household disposable income for the even ones. Compared to the still credit constrained sample, the credit unconstrained sample will decrease the household saving rate by 11 to 27 percent, that is to say, the average "treatment" effect of credit unconstrained on the household saving rate is impressive, and credit constraint is an anchor to reduce the saving rate especially for poor households. Columns 3 to 6 of table B.6 show the estimated effects of credit unconstrained on the household saving rate is negative significant by using the financial credit measure (Zeldes, 1989a) by using the CHFS and CFPS data. The most recent CHIP data sample of the year

2013 is in the year 2009, and it should be more problematic due to the longer time span. The Difference-in-Difference (DID) specification and the t-test for the two sample groups are also the same The magnitude and significance of the interaction term are similar for the financial constraint measures. Compared to the still financial credit constrained sample, the financial credit unconstrained sample will decrease the household saving rate by 21 to 26 percent and 24 to 26 percent for CHFS and CFPS data, respectively. The conspicuous effect of the borrowing and financially unconstrained on the household saving rate confirm our data facts and theoretical implication that the existence of the liquidity constraint leads to a significantly higher household saving rate and also expect the liquidity constraint lead to higher aggregate saving rate if we measure household sample as more aggregate level.

3.5.3 Income inequality and saving rate

Evidence across counties: cross-section regressions across county We have seen that the household saving rate response to a move up in the current and permanent income quintile positive significantly. We address the research question what the general equilibrium effect of the aggregate household saving rate from a rise in the income inequality. We raised this by examining the calculated county-level data and indicators from CHFS 2015, CFPS 2014 and CHIP 2013 datasets for any association between local total saving rate and Gini coefficient. The three dataset cover 353 counties, 334 counties and 212 counties, respectively. The empirical exercise is valid for the causal relationship between the local total saving rate and inequality measure because the inequality across the county has no association with the county local saving rate after controlling for the location fixed effects. The income inequality indicators we use is the Gini coefficient and other measures such as coefficient of variation, Theil index, Mehran index, Piesch index, Kakwani index and the income ratio of the top 20% of households to the bottom 20% of households or relative mean deviation (Chu and Wen, 2017; De Maio, 2007). We perform the cross-sectional regression that links a county i's saving rate in the year 2013-2015 with the inequality index, controlling for location fixed effects and other factors. The variable we controlled maintain the previous specification as the (Schmidt-Hebbel and Serven, 2000) such as the local level young-dependent ratio (`YDratio`) and the old-dependent ratio (`ODratio`). To be precise, we report the results both controlling the county level young-dependent ratio (`YDratio`) and the old-dependent ratio (`ODratio`) or not to check the robustness of the effect. The location local saving rate and the inequality indexes, along with the young-dependent ratio (`YDratio`) and the old-dependent ratio (`ODratio`) are computed from CHFS 2015, CFPS 2014 and CHIP 2013 datasets. The saving rate of each county is defined as local total income minus consumption, divided by income. The inequality measure is defined as (De Maio, 2007) and local per-capita disposable income is defined as the county local total income divided by the local total population. We cluster the standard errors by province. **Income inequality and aggregate household saving rate** The odd column of table B.7 report the regression results with only county per-capita income and Gini coefficient as the regressors, using the CHFS, CFPS and CHIP dataset, respectively. The effect of income inequality on local saving rate is significantly positive at 5 percent level: a 1 percent increase in the Gini coefficient is associated with higher county saving rate by 0.18-0.20 percent points. The coefficient on local per-capita income is around 0.33 and statistically different from zero at the 1 percent level. In other words, CHFS 2015 data reveals that the county level saving rate tends to be higher in the county with more unbalanced income distribution. In any case, we note that similar household saving rate age-profile patterns relative to the prediction of the life cycle hypothesis are documented in Chamon and Prasad (2010) since we note that the share of young-age population and the old-age population has a positive significant coefficient. A 1 percent increase in the the old-dependent ratio (`ODratio`) or young-dependent ratio (`YDratio`) is associated with 0.35 to 0.6 percent higher county saving rate. These results may imply that that old-age households or household with children tend to save more than working-age household, however the significant association appears in the CHFS dataset is not consistent for other two datasets. Columns 3 to 6 of table B.7 show the estimated effects of Gini coefficient on local saving rate is not significant: a 1 percent increase in the Gini coefficient is associated with insignificant higher local saving rate by 0.13 and 0.02 percent points for CFPS and CHIP data, respectively. To save space, we do not report the results using other inequality measures, and some of the association between the income inequality and local saving rate are significantly positive at 5 to 10 percent level. To summarize, although the positive relationship between high-income inequality and a high county level saving rate is not robust and statistically significant, we can conclude that the increase in the current income inequality would make the local total household saving rate even higher.

3.5.4 Marginal propensity to consume (MPC)

Friedman's permanent income hypothesis (PIH) states that one's current consumption is determined not just only by the current income but also by the expected future income (permanent income). The hypothesis suggests that the changes in consumption are mainly driven by the changes in permanent income rather than the changes in current income, which implies that the marginal propensity to consume (MPC) out of permanent income would be greater than the MPC out of current income. To know how Chinese household would respond to the changes in current income, we estimate the empirical model (3.12) by running five separate cross-sectional regressions, with each one focusing on one income quintile sample. Table B.8 presents the estimates of the MPC out of each type of income. In panel (a) of Table B.8, for both types of income, the estimated MPCs are significantly positive in the CHFS. The estimated MPCs out of permanent income range from 223 RMB per 1000 RMB for the highest-income quintile households to 778 RMB per 1000 RMB for the lowest. We notice that the coefficients are decreasing with income classes except for the quintile 4. For the quintile 4, the coefficients are 541 RMB per 1000 RMB, which is not only greater than the quintile 5 but also than the quintile 3. The estimated MPCs out of current income share the similar pattern to the permanent income. The lowest-income quintile has the largest coefficient of 788 RMB per 1000 RMB, and the highest-income quintile has the smallest MPC with 178 RMB per 1000 RMB. Except for the quintile 4, the coefficients exhibit a decreasing pattern with as income quintiles move up. The estimated MPC out of current income for the quintile 4 is 507 RMB per 1000 RMB, which is larger than the quintile 3 and 5. These results indicate that the MPC out of both types of income for the bottom 20 percent households is much higher than the top 20 percent.

Panel (b) shows the results from similar regression using CFPS. The MPCs for both types of income are highly significant. The estimates of the MPCs out of current income range from 336 RMB per 1000 RMB for the quintile 5 to 927 RMB per 1000 RMB for the quintile 1. They display a decreasing pattern with income levels except for the quintile 2, whose estimated MPC is 361 RMB per 1000 RMB, but it is still larger than the quintile 5. For the MPC out of permanent income, excluding the quintile 2, the coefficients are still exhibiting a decreasing pattern with income classes. A little bit differences from the MPC out of current income is that the coefficient for the quintile 2 is 347 RMB per 1000 RMB, which is the smallest among the five quintiles. The MPC out of both types of income for the lowest-income quintile is again much higher than the highest-income quintile.

The results of estimation from CHIP are shown in panel (c) of the table. The estimated MPCs out of permanent income is 613 RMB per 1000 RMB for the bottom 20 percent households. It increases a little bit to 623 RMB per 1000 RMB for the second income quintile and then decreases steadily from 510 RMB per 1000 RMB for the third income quintile to 233 RBM per 1000 RMB for the top 20 percent households. All the numbers are significantly positive. For the estimated MPCs out of current income, the CHIP data exhibit a diminishing pattern as income quintiles move up, decreasing from 563 RMB per 1000 RMB for the bottom 20 percent households to 167 RMB per 1000 RMB for the top 20 percent households. Except for the coefficient for the top 20 percent households, all others are statistically significant positive. Besides, the results in the CHIP exhibit the pattern implied by the permanent income hypothesis (PIH) that the coefficients for the permanent income higher than that for the current income.

In sum, although we do not see a diminishing MPC with income classes across three data sets, there is still a not surprising pattern that the MPC out of both types of income for the bottom 20 percent households are much higher than that for the top 20 percent households. The results are consistent with empirical evidence in the literature.

3.6 Conclusion

We show in this paper the role of income inequality interacting with liquidity constraints in explaining the high household saving rate in China. In a simple two-period model, households are heterogeneous in income and subjective discount factor, and whether the liquidity constraint is binding, consumption and saving rate are endogenously determined. The model generates several predictions consistent data facts: (1) the rich save more; (2) the proportion of constrained households for the poor is higher than that for the rich; (3) liquidity constraints would increase household saving rate. (4) when income inequality increases, the rich save even more, in the meanwhile, the poor would also save more due to the binding liquidity constraints, and thus the aggregate household saving rate would rise.

Using three sources of large, nationally representative household survey data, the China Household Finance Survey (CHFS), the China Family Panel Studies (CFPS), and the Chinese Household Income Project (CHIP), we provide direct empirical evidence implied by the theoretical model. We find that in China, (1) the top 20 percent permanent income households' saving rate is 19–23 percent significant higher than the bottom 20 percent households'. (2) the bottom 20 percent permanent income households are more likely to face a borrowing constraint, with a 12–20 percent significant higher probability; (3) the existence of liquidity constraints would lead to a significant increase of more than 20 percent in the household saving rate; (4) income inequality would have a significant positive impact on the aggregate household saving rate ath the county level, with a 1 point on a scale of 100 measure increase in the Gini coefficient leading to an increase of 0.2 percent in the aggregate saving rate; (5) the estimated MPC for the top 20 percent households range from 200 to 400 RMB per 1000 RMB, while for the bottom 20 percent households, the range from 600 to 900 RMB per 1000 RMB.

These findings would have significant policy implications. The Chinese governmentâĂŹs policies on reducing the saving rate have not yet produced substantial results. If income inequality and liquidity constraints were the key reasons for the high aggregate household saving rate, the resulting policy would be drastically different. For example, it is appropriate for the Chinese government to design some income redistribution programs (such as EITC) to reduce income inequality or devote more resources to support the credit market development. An economic policy of tackling income inequality would lower the aggregate saving rate, thus becoming a policy of economic transition and growth.

4. DOES THE COLLEGE ADMISSION POLICY IMPROVE HIGH SCHOOL EDUCATION IN CHINA'S POOR COUNTIES?

4.1 Introduction

In the fourth chapter, we apply causal inference methodology to evaluate and identify the effect of the "Program of College Admission for Poor Counties" on high school education in current China. In view of educational disadvantages of students from poverty regions with regard to unequal access to college, Ministry of Education, along with other four ministries, initiated a "Program of College Admission for Poor Counties"[1], a policy that enhance college admission upon China's "National College Entrance Exam" (*Gaokao*) for high schools' graduates in poverty-stricken counties (henceforth "poor counties")[2] from the year 2012. It assigns exclusive admission quotas to admit high school graduates from designated poor counties. The program is motivated to upgrade the development capacity of the poverty regions and expand the college admission to students from those areas, thereby promote the education equality. We want to empirically evaluate the impact of this college admission policy on "upstream" high school education since one of the policy objectives is to incentivize the development of pre-tertiary education in those poor counties.

In this study, we examine the effect of this college admission policy on high school education performance in poor counties in Gansu, a northwestern province in China. We focus on the senior high school entry rate, and enrollment rate before and after the policy takes effect by applying a difference-in-differences (DD) empirical strategy. We do not look into the policy effect on education before senior high school because primary school and junior high school education constitute the "Nine-Year Compulsory Education" by law in China. Our

[1]Ministry of Education of the People's Republic of China (2012) "Program of College Admission for Poor Counties". http://www.moe.gov.cn/srcsite/A15/s7063/201203/t20120319_134392.html

[2]The list of "poor counties" in China is released by the State Council Leading Group Office of Poverty Alleviation and Development. The list was first released in 1986 and then revised in 1994, 2006 and 2012. The "Program of College Admission for Poor Counties" is based on the most recent version (2012). The State Council Leading Group Office of Poverty Alleviation and Development: http://www.cpad.gov.cn/art/2012/6/14/art_50_23717.html

results show that policy is effective in terms of significant increase in entry and enrollment rate. In particular, this study sheds light on the "general equilibrium effect" of this education policy, which aims to promote access to college, actually improves the schooling for "upstream" high school education in those policy treated counties. Our paper contributes to the literature by showing that policy for mitigating educational inequality generates "general equilibrium effect" that arises at high school stage through the incentives for schooling.

Roadmap The rest of the paper is organized as follows. In section 4.2, we introduce the background of the admission program "Program of College Admission for Poor Counties " and other related education admission policy and literature. Section 4.3 presents the data using for the empirical analysis. Section 4.4 describes the empirical methodology. The results and analysis are presented in Section 4.5. Section 4.6 concludes. The Appendix contains proofs, figures, tables, and other details.

4.2 Background

Education System in China In China, the state plays a central role in shaping the educational system and distributing educational opportunities. Major shifts in policies on education have dramatically altered individual's life chances during the pre-reform era from the 1950s to the 1970s (Wu and Zhang, 2010; Deng and Treiman, 1997; Zhou, Moen, and Tuma, 1998). The economic reform and open-up policy since the late 1970s generate rapid economic growth, as well as increasing demand for more skilled labor and more talented personnel. Education has become an increasingly important factor in determining individual's socioeconomic attainment (Wu and Zhang, 2010). A new wave of policy updates addressed the goal of universal compulsory education. In 1985, *the Decision on the Reform of the Education System* was initiated, followed by *the Compulsory Education Law* in 1986. By the mid-1990s, these goals were achieved by and large. However, the expansion of education beyond the compulsory level had been slower until the expansion of higher education from the late 1990s. While the expansion of senior high school education did not keep the pace with the higher education level, which may have significant consequences on educational inequality.

Unequal Access to College Education inequality is a major concern on the policy agenda in China as well as in the rest of the world. In particular, a growing debate has been addressing the inequality in access to higher education. Discussions on the unequal access to college in China stem from the perspective of rural-urban segmentation, as well as variation across regions, particularly for those areas of poverty (Hannum, 1999, 2003; Hannum and Wang, 2006; Tsang, 1996, 2000; Wu, 2010; Wu and Zhang, 2010; Zhang and Kanbur, 2005; Zhou *et al.*, 1998). Even after mass college expansion started in 1998, rural youth from poor counties were still much less likely to access college than urban students (Li *et al.*, 2015).

Incentive for Schooling and Education Decisions Incentive for households to send their children to school depends on the expected return to education, which can be decomposed into likelihood of access to education (opportunity) and return to education, as well as those constraints, such as institutional barriers, opportunity cost, and financial constraint (Brown and Park, 2002; Liu, Zhang, Luo, Wang, Rozelle, Sharbono, Adams, Shi, Yue, Li, et al., 2011).

Unequal access to education with respect to student's family background, the location of origin has been increasing over recent decades (Li, Loyalka, Rozelle, Wu, and Xie, 2015; Zhang and Kanbur, 2005; Hannum and Meiyan, 2006) Loyalka, Chu, Wei, Johnson, and Reniker (2017) examine inequalities emerge along the pathway to college in China. They find that the most significant inequalities in college access arise when students move from junior high to senior high school. The Key-school system plays an essential role in the pathway to college in China (Ye, 2015). Wang, Liu, Zhang, Luo, Glauben, Shi, Rozelle, and Sharbono (2011) find that the college matriculation rate of the poor is substantially lower than students from non-poor families. The real hurdles keeping the rural poor from pursuing a college education arise as early as preschool and elementary school years and persist throughout the entire schooling system. Also, there is supply-side intervention on quality of teaching and schools. Loyalka, Wei, Song, Zhong, and Chu (2014) conduct an impact evaluation of the "Innovative High Schools" intervention on the college admission outcomes of students in the poor counties of northwest China. Findings show that this intervention increases the chances that a typical student from a poor county will gain admission to non-elite college, but not elite colleges.

Conditional on access to education, return to education, particularly on higher education, further shape the incentive for schooling. Zhang, Zhao, Park, and Song (2005) finds a dramatic increase in the returns to education in urban China, from only 4 percent per year of school in 1988 to 10.2 percent in 2001. In particular, most of the rise in the returns to education occurred after 1992 and reflected an increase in the wage premium for higher education. Li, Liu, and Zhang (2012) use data on twins to estimate the return to education in urban China. The findings show that high school education may mainly serve as a mechanism to select college students, but as a human capital investment per se, it has low returns in terms of earnings. Meng, Shen, and Xue (2013) examine the causes of the increase in earnings inequality of Chinese urban male workers. The slowing down of the reward to skills is, to some extent, due to the college expansion that occurred at the end of the 1990s. Jia and Li (2016) find a sizable wage premium of elite education. However, access to elite education does not promote one's entry into the elite class, but parents' elite class does. Also, they find that access to elite education does not alter the intergenerational link between parents' status and children's status.

Another strand of literature investigates the cause of individual and household educational decision with respect to certain outcomes such as enrollment and dropout. Liu, Zhang, Luo, Wang, Rozelle, Sharbono, Adams, Shi, Yue, Li, et al. (2011) find that early commitment on financial aid helps the student to make less distorting college decisions. Loyalka, Liu, Song, Yi, Huang, Wei, Zhang, Shi, Chu, and Rozelle (2013) study the effects of providing information on returns or career planning skills on student dropout, academic achievement and plans to go to high school. The findings show that information does not have significant effects on student outcomes, but counseling increases dropouts and seems to lower academic achievement. Mo, Zhang, Yi, Luo, Rozelle, and Brinton (2013) examine the effectiveness of a Conditional Cash Transfer (CCT) program on the dropout rate. The results show that the program reduces dropout by 60 percent. In particular, the program is most effective among students with poor academic performance.

Affirmative Action In China, the government initiated the affirmative action policies from the mid-1980s, which offers preferential treatment to ethnic minorities. For example, minority students receive bonus points on "National College Entrance Exam", thereby get more chance for admission (Sautman, 1998; Zhou and Hill, 2009). While in the United States, several states eliminated affirmative action admissions policies for public universities during the late 1990s. Some of these states substituted a program that grants admission to the top x% of each high school's graduating class. These new programs were motivated to restore minority college enrollments to previous levels. There is a growing debate about the effect of the x% rule. Dickson (2006) finds that the end of affirmative action significantly lowered the percentage of Hispanic and black students applying to college in Texas. A year after the end of affirmative action in Texas, the Texas State Legislature passed a percent plan that guarantees students who graduate in the top 10 percent of their high school class admission to any public college in Texas. The percent of minority students applying to college increased significantly when the plan was accompanied by changes in financial aid. Cullen, Long, and Reback (2013) find unintended consequences of this policy that aims to improve college access for disadvantaged and minority students. This policy encourages strategic high school enrollment that might induce would-be eligible students to choose to attend lower-achieving schools than they otherwise would. Long (2004) finds that x% programs are unable to replace traditional affirmative action and maintain the share of minority students.

" Program of College Admission for Poor Counties" "Program of College Admission for Poor Counties" is an important part of the grand strategy of poverty alleviation, as well as reform on college admission scheme. Specifically, it imposed 10,000 exclusive admission quotas to first-tier colleges to admit high school graduates from poor counties nationwide.

The quotas are increased to 30,000 in 2013, 50,000 in 2014 and 2015, 60,000 in 2016 and 63,000 in 2017. It alters incentive for schooling for students in poor counties in important ways. First, students who might have been otherwise hopeless for college access now become more promising. Some of them who might not choose to go to senior high school now choose to be enrolled. Also, those who might drop out during senior high school now choose to stay. So, this policy increases the opportunity cost of being out of high school education. In addition, this policy reduces the opportunity cost for senior high schools in poor counties preparing their students for the college education when facing expanded college access opportunities.

We want to examine empirically whether people respond to this policy in terms of choices such as to be enrolled in high school. This paper estimates the impact of the admission program on high school education in poor counties using data from Gansu, a northwestern province in China. Our study is the first empirical evaluation for this policy with respect to the policy object of improving pre-tertiary education in poor counties.

4.3 Data

This study uses county-level data of Gansu province from 2005 to 2015 from the *Gansu Statistical Yearbook*, China Population Census 2010 and Population Mini Census 2005. This is a balanced panel data on 86 counties spanning 11 years. Gansu, located in the northwest of China, is a province with relatively lower economic advancement. The GDP per capita in Gansu in 2015 is, compared with average nationwide. The province includes 14 prefectures and 86 counties. The population is 2.6 million and *per–capita* GDP 27,458 RMB as of the 2016.[3] The key high schools in public school system are unevenly distributed across regions within the province, as shown in Figure C.2 at prefecture level. Figure C.3 shows the great variation of government expenditure per Student on Education by different prefecture cities in Gansu Province.

The outcome variables of our study include a collection of indicators from senior high

[3]Gansu Province Bureau of Statistics http://www.gstj.gov.cn/www/HdClsContentDisp.asp?Id=34639

school education statistics, namely, entry rate and enrollment rate. All of the schools, teachers, and students are referred to those in *senior high school* (henceforth "high school"). The data of the outcomes variables are documented at the county level. The construction of the indicators of educational outcomes follows Ministry of Education (2015). Specifically, the outcome variables are defined as follows:

Entry rate 1: Entrants per 100,000 population

$= \text{Entrants} \div \text{Total Population} \times 100{,}000$

Enrollment rate 1: Enrollments per 100,000 population

$= \text{Enrollments} \div \text{Total Population} \times 100{,}000$

Graduation rate 1: Graduates per 100,000 population

$= \text{Graduates} \div \text{Total Population} \times 100{,}000$

In addition, when checking the robustness of our findings, we refer to alternative definitions of educational outcomes from OECD and UNESCO, which are defined as follows:

Entry rate 2 $=$ Entrants $\div$ Entrant-Age Population (Age 15)

Enrollment Rate 2 $=$ Enrollments $\div$ Enrollment-Age Population (Age 15-17)

Graduation rate 2 $=$ Graduates $\div$ Graduation-Age Population (Age 17)

High school graduates from 58 designated poor counties, which accounts for about two-thirds (68%) of all 86 counties in Gansu, are eligible for the "Program of College Admission for Poor Counties". Summary of statistics for outcome variables is shown in Table C.1. In addition, we control for other factors that may affect high school education, such as the number of teachers, education expenditure in public finance budgetary expenditure, and GDP *per capita*.

4.4 Empirical Strategy

Research Design and Identification Assumption This article explores the within-county variations to identify the effect of "Program of College Admission for Poor Counties", which is targeting students from poor counties, on the schooling in senior high school stage.

To distinguish the effect of the policy from confounding factors, we exploit the within-country variation by applying a difference-in-differences (DD) identification strategy. We compare the within-county changes in schooling outcomes of policy-eligible counties before and after 2012 when the policy takes effect, to the within-county changes in those non-eligible counties over the same period. We ask the question whether outcomes change more in policy counties than non-policy counties. The key identification assumption is that in the absence of the policy, those policy counties would have experienced changes in schooling outcomes similar to non-policy counties.

The data allow us to test and relax this assumption in several ways. First, using graphical evidence and regression results, we can show whether the outcomes of the treatment and control counties did not diverge before the policy takes effect. In addition, we can check whether our findings are robust to the inclusion of time-varying covariates such as GDP *per capita*, education expenditure in public finance, and the number of teachers. This will reveal whether other known determinants of schooling outcomes were orthogonal to the treatment of the policy.

Baseline Model Specification We estimate the following baseline regression equation:

$$Y_{ijt} = \beta_i + \eta_{jt} + \lambda \times \text{Policy}_{ijt} \times \text{After}_{ijt} + \mu \times \mathbf{X}_{ijt} + \epsilon_{ijt} \tag{4.1}$$

In Equation B.1, subscript i indicates a county, j is the prefecture that the county i is affiliated, and t denotes the year of observation. Yijt is the high school education outcome, i.e., entry rate and enrollment rate. There are two key indicator variables (0/1 dummy) in this specification: Policyijt equals one if county i is eligible for the college admission policy; Afterijt indicates that it occurred after the policy is initiated. The vector **X**ijt of controls includes educational expenditure as a share of total public expenditure and number of teachers (in logarithm). Also, we control all time-invariant or fixed county characteristics in the vector β_i. In addition, we include a prefecture-year fixed effects, η_{jt}. Robust standard errors are clustered at prefecture level since the sample is limited to Gansu province and the unit is county level. The coefficient of primary interest is λ, which is a difference-in-differences estimator of the impact of college admission policy on high school education outcomes. It measures the change in outcomes after the policy is initiated, relative to before, among policy counties to non-policy counties.

Dynamic Effects and Test for Parallel Trends Assumption The underlying assumption for an unbiased estimate of λ is that in the absence of the admission policy of exclusive quotas, poor counties would have experienced changes in high school's performance similar to non-poor counties. The sample data allow us to test this identifying assumption. We do a formal test by including indicators of years and their interaction with the policy in the regression. We want to know whether counties that are eligible for the policy diverge from non-policy counties before 2012. If they do, as shown in the estimated coefficients of the years before 2012 interacting with policy indicator, then it suggests that the identifying assumption of our empirical approach is violated. In addition, the policy quotas have been increasing since 2012, so we want to capture the dynamic effects of this program in the expansion. We estimate the following equation 4.2 of the dynamic effects by year:

$$Y_{ijt} = \beta_i + \eta_{jt} + \Sigma_t \lambda_t \times \text{Policy}_{ijt} \times \text{Year}_t + \mu \times \mathbf{X}_{ijt} + \epsilon_{ijt} \tag{4.2}$$

The parallel-trend assumption requires that λ_t should not significantly differ from zero for any year before 2012.

Propensity Score Matching (PSM) Since the placement of the policy is not necessarily random, then we want to address the confounding factors for identifying the policy effect. A particular concern here is that the initial conditions are likely to influence policy placement and outcome of interests as well. If so, then the DD method will generate biased estimation. We construct the control and treatment group of counties in our research design using matching method (Rosenbaum and Rubin, 1983; Hirano, Imbens, and Ridder, 2003). Since policy (treatment) counties and non-policy (control) ones in the same areas are supposed to share similar characteristics, then we use the matching method to construct the counterfactual for the policy counties. Propensity score matching (PSM) method is used here for balancing treatment and control counties in terms of initial conditions, accommodating the possibility of observable time variant selection bias due to initial conditions. We predict the propensity score, i.e., the probability of being selected into policy, for each county using population size, GDP *per capita*, fiscal revenue *per capita*, rural residents net income *per capita*, non-agricultural population (as a ratio of total population), prefecture dummy (indicating which prefecture a county belongs to), and those senior high school education outcomes in years prior to policy. Since the "poor counties" list is drafted according to the indicators that are closely relevant to poverty during 2007 to 2009[4], we use the same period for predicting propensity score on initial conditions.

Placebo Test To justify our identification assumption, we examine whether there are "placebo" effects as if the years before 2012 are starting year of the college admission program. We use the year 2009 as a placebo cutoff for the timing when the policy is initiated. It is expected that outcomes after the year 2009 should be unaffected by the college admission policy, which will provide further evidence on whether pretreatment trends are parallel in poor and non-poor counties.

4.5 Results and Discussion

Main Results Table C.2a reports the results for entry rate, i.e. number of high school entrants per 100,000 population, and Table C.2b for enrollment rate, Table **??** and **??** for enrollment rate both from Equation (1). Panel A includes estimated results from Equation (1) without PSM, while results with PSM are reported in Panel B. Column (1) and (3) in the tables are results from simple DD specification without controls, while Column (2) and (4) with controls. The policy effect is significant positive, increasing by 99- 224 entrants,

[4]The State Council Leading Group Office of Poverty Alleviation and Development: http://www.cpad.gov.cn/art/2012/6/14/art_50_23717.html

and 317-586 enrollments per 100,000 population in those poor counties.

Robustness Checks To check the robustness of our findings on policy effect, we examine the alternative definition of those outcomes aforementioned. The entry rate is measured by the ratio of entrants over the age-cohort population, likewise enrollment rate. Table C.2c and C.2d also report the results for alternative outcome measurements, which are consistent with the results above. The entry rate has been enhanced by 1.3-7.6%. The enrollment rate has been increased by 1.2-7.3%.

Placebo Test Furthermore, we conduct a placebo test to see whether a placebo year of starting the policy will falsify our results. We use the year of 2009 as the start of the policy to examine the change in outcomes from 2006 to 2011. The results shown in Table 6 turn out no policy effect upon a placebo year of starting the policy.

Dynamic Effects and Test for Parallel Trends Assumption Table C.7 present the results for the dynamic effects of the admission policy estimated from Equation (2). Alternatively, Figure 1 displays the dynamic effect graphically. When the policy effect is pinned down to each year, we can see that before 2012, the policy effect is not significantly different from zero, confirming the parallel-trend assumption. The policy effect from 2012 onward is significant for the year 2012 and 2015, partly reflecting the changing size of policy quotas. Also, people take time to learn the policy and integrate it into the education decisions.

Heterogeneous effect, Alternative Explanations, and Underlying Mechanism We also report the heterogeneous effect of the admission policy on the outcomes for the subsample grouped by the per-capita GDP in table C.5a to table C.5e. We explore two alternative explanations that can drive the variation of the difference between policy-eligible counties and non-policy counties. First, there is a supply-side driven quality improvement in high school education in poor counties relative to non-poor counties. Second, there is a possibility that some students from non-poor counties strategically transfer to poor counties to take advantage of the college admission policy. Both can bring about selection bias to our estimation.

Education Quality We choose two proxies for supply-side driven quality improvement in high schools: student-teacher ratio and education expenditure in county's public finance. We test whether the college admission policy for poor counties affects county fiscal expenditure on education or student-teacher ratio, using the regression specification of Equation (1). The results are reported in Table C.6a and Table C.6b. All of the estimated coefficients are not statistically significant except for the education expenditure regression without controls or PSM. So, it is unlikely that the estimated effect of the college admission policy on high school education is not driven by local supply-side driven quality improvement in high schools.

Migration There is also a concern for student migration, i.e., transfer between high schools between policy and non-policy counties, in response to the policy. The identifying assumption would be violated if the policy causes students with systematically different unobserved abilities to transfer from non-policy counties to policy counties. We could not directly test this concern since the data unit is aggregated at the county level. However, the policy is implemented strictly, which virtually rules out opportunities for students to transfer strategically. Also, we check the migration as a share of the cohort population aged 15 from non-policy counties to policy counties, which is -1.6%, a magnitude that not sizable enough to counteract the policy effect.

In sum, the results show that this policy significantly increased senior high school entrants and enrollments in those poor counties in Gansu from 2012 to 2015. Also, we can see that the policy takes time to be effective, as students and their families, as well as schools, take time to learn the policy and therefore integrate it into their schooling decision-making. Collectively, these findings suggest that county-level aggregate schooling at senior high school stage is effectively enhanced by this college admission policy.

4.6 Conclusion

This article uses difference-in-differences strategy to estimate the effect of a college admission policy on the senior high school education. This policy, the "Program of College Admission for Poor Counties", assigns exclusive admission quotas to enroll students from poor counties. We find that schooling at senior high school stage is effectively improved in terms of entry and enrollment. Our findings also have significant policy implications. This policy for addressing unequal access to college education in fact enhanced schooling at high school stage, which exhibits a "general equilibrium effect", as senior high school and college are connected in educational continuation decisions. This finding indicates that the policy effect meets one of the policy objectives, that is, it provides positive incentives for promoting pre-tertiary education.

5. CONCLUSION

To conclude, the main focus of my doctoral research is to apply causal inference technique to identify the effects of unbalanced local sex ratio and income distribution, liquidity constraint on saving rate of household sector and, in addition, to apply causal inference methodology to evaluate and identify the effect of the "Program of College Admission for Poor Counties" on high school education in current China.

In the first essay, we answer this question "Can the sex ratio explain high China's household saving rate today?". The high aggregate household saving rate is one of the unique features of the Chinese economy. Over the period of China rapid income growth, China's sex ratio, which is defined as the ratio of the number of boys at birth to the number of girls at birth, has been also experienced a rapid a growth. The common trend has captured economists' attention on linking the saving rate to the sex ratio. Can the sex ratio explain high China's household saving rate? Starting with Wei and Zhang (2011a), they argue a explanation called "competitive saving motive" that increasingly unbalanced premarital sex ratio requires both rural and urban son-families to save increasingly more to compete in the marriage market and show that the rapid increase in premarital sex ratio during 1990-2007 can potentially explain about half of the sharp rise in China's household saving rate.

In this paper, we reexamine the impact of the sex ratio on the household saving rate. First, we use the CHIP data — the same sources of data as used by Wei and Zhang (2011a) — in an attempt to replicate and extend the estimates reported in Wei and Zhang (2011a). We use local sex ratio inferred from 2000 China population census and find the competitive saving motive only holds for the household in rich counties. The effects vanish for CHIP 2002 full sample with census data in 2000, especially for poor households and households in poor counties. We argue that unbalanced migration of premarital male and female is the major explanation for the vanishing effect.

Next, we turn to another nationally representative data set China Household Finance

Survey (CHFS) 2015 wave for estimating the effect of the sex ratio on the household saving rate. The result indicates that the competitive saving motive still holds, but only in the rural area. Specifically, in the rural, the effects on households with a son is much smaller, and the effects on households with a daughter are also significant positive. However, there are no effects on urban households.

Finally, we use China population census and provincial panel data as the same data sources as Wei and Zhang (2011a) used, and adopt a more precise algorithm to compute the sex ratios of the 31 provinces for 7 years (1985, 1990, 1995, 2000, 2005, 2010, and 2015). Our results report that the competitive saving motive exists, but only in the rural area. By estimation and computation, an increase in sex ratio from 1985 to 2015 can explain about 28% of the actual increase of the increase of rural saving rate.

In the second essay, we show in this paper the role of income inequality interacting with liquidity constraints in explaining the high household saving rate in China. In a simple two-period model, households are heterogeneous in income and in subjective discount factor, and whether the liquidity constraint is binding, consumption and saving rate are endogenously determined. The model generates several predictions consistent data facts: (1) the rich save more; (2) the proportion of constrained households for the poor is higher than that for the rich; (3) liquidity constraints would increase household saving rate. (4) when income inequality increases, the rich save even more, in the meanwhile, the poor would also save more due to the binding liquidity constraints, and thus the aggregate household saving rate would rise.

Using three sources of large, nationally representative household survey data, the China Household Finance Survey (CHFS), the China Family Panel Studies (CFPS), and the Chinese Household Income Project (CHIP), we provide direct empirical evidence implied by the theoretical model. We find that in China, (1) the top 20 percent permanent income householdsâĂŹ saving rate is 19–23 percent signiïňĄcant higher than the bottom 20 percent householdsâĂŹ. (2) the bottom 20 percent permanent income households are more likely to face a borrowing constraint, with a 12–20 percent signiïňĄcant higher probability; (3) the existence of liquidity constraints would lead to a significant increase of more than 20 percent in the household saving rate; (4) income inequality would have a positive significant impact on the aggregate household saving rate ath the county level, with a 1 point on a scale of 100 measure increase in the Gini coefficient leading to an increase of 0.2 percent in the aggregate saving rate; (5) the estimated MPC for the top 20 percent households range from 200 to 400 RMB per 1000 RMB, while for the bottom 20 percent households, the range from 600 to 900 RMB per 1000 RMB.

These findings would have important policy implications. The Chinese governmentâĂŹs policies on reducing the saving rate have not yet produced substantial results. If income inequality and liquidity constraints were the key reasons for the high aggregate household saving rate, the resulting policy would be drastically different. For example, it is appropriate for the Chinese government to design some income redistribution programs (such as EITC) to reduce income inequality or devote more resources to support the credit market development. An economic policy of tackling income inequality would lower the aggregate saving rate, thus becoming a policy of economic transition and growth.

The third essay uses difference-in-differences strategy to estimate the effect of a college admission policy on the senior high school education. This policy, the "Program of College Admission for Poor Counties", assigns exclusive admission quotas to enroll students from poor counties. We find that schooling at senior high school stage is effectively improved in terms of entry and enrollment. Our findings also have significant policy implications. This policy for addressing unequal access to college education in fact enhanced schooling at high school stage, which exhibits a "general equilibrium effect", as senior high school and college are connected in educational continuation decisions. This finding indicates that the policy effect meets one of the policy objectives, that is, it provides positive incentives for promoting pre-tertiary education.

REFERENCES

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2017): "Income and wealth distribution in macroeconomics: A continuous-time approach," Discussion paper, National Bureau of Economic Research.

Aiyagari, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109(3), 659–684.

Banerjee, A., X. Meng, T. Porzio, and N. Qian (2014): "Aggregate Fertility and Household Savings: A General Equilibrium Analysis using Micro Data," Discussion paper, National Bureau of Economic Research.

Bhalla, S. S. (1980): "The measurement of permanent income and its application to savings behavior," *Journal of Political Economy*, 88(4), 722–744.

Böhlmark, A., and M. J. Lindquist (2006): "Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden," *Journal of Labor Economics*, 24(4), 879–896.

Brown, P. H., and A. Park (2002): "Education and poverty in rural China," *Economics of Education Review*, 21(6), 523–541.

Carroll, C. D., and M. S. Kimball (1996): "On the concavity of the consumption function," *Econometrica: Journal of the Econometric Society*, pp. 981–992.

Carroll, C. D., and D. N. Weil (1994): "Saving and growth: a reinterpretation," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 40, pp. 133–192. Elsevier.

Chamon, M. D., and E. S. Prasad (2010): "Why are saving rates of urban households in China rising?," *American Economic Journal: Macroeconomics*, pp. 93–130.

Chang, S., and X. Zhang (2012): "The economic consequences of excess men," Discussion paper, IFPRI Discussion Paper.

——— (2015): "Mating competition and entrepreneurship," *Journal of Economic Behavior & Organization*, 116, 292–309.

Choukhmane, T., N. Coeurdacier, and K. Jin (2013): "The one-child policy and household savings," .

Chu, T., and Q. Wen (2017): "Can income inequality explain China's saving puzzle?," *International Review of Economics & Finance*, 52, 222–235.

Cooper, R., and G. Zhu (2017): "Household Finance in China," Discussion paper, National Bureau of Economic Research.

Cullen, J. B., M. C. Long, and R. Reback (2013): "Jockeying for position: Strategic high school choice under Texas' top ten percent plan," *Journal of Public Economics*, 97, 32–48.

Curtis, C. C., S. Lugauer, and N. C. Mark (2015): "Demographic patterns and household saving in China," *American Economic Journal: Macroeconomics*, 7(2), 58–94.

De Maio, F. G. (2007): "Income inequality measures," *Journal of Epidemiology & Community Health*, 61(10), 849–852.

Deng, Z., and D. J. Treiman (1997): "The impact of the cultural revolution on trends in educational attainment in the People's Republic of China," *American Journal of Sociology*, 103(2), 391–428.

Dickson, L. M. (2006): "Does ending affirmative action in college admissions lower the percent of minority students applying to college?," *Economics of Education Review*, 25(1), 109–119.

Ding, W., and Y. Zhang (2014): "When a son is born: The impact of fertility patterns on family finance in rural China," *China Economic Review*, 30, 192–208.

Du, Q., and S.-J. Wei (2013): "A theory of the competitive saving motive," *Journal of International Economics*, 91(2), 275–289.

——— (2016): "A Darwinian perspective on exchange rate undervaluation," *European Economic Review*, 83, 111–138.

Dynan, K. E., J. Skinner, and S. P. Zeldes (2004): "Do the rich save more?," *Journal of Political Economy*, 112(2), 397–444.

Edlund, L., H. Li, J. Yi, and J. Zhang (2013): "Sex ratios and crime: Evidence from China," *Review of Economics and Statistics*, 95(5), 1520–1534.

Friedman, M. (1957): "The permanent income hypothesis," in *A theory of the consumption function*, pp. 20–37. Princeton University Press.

Fuchs-Schündeln, N., and M. Schündeln (2005): "Precautionary savings and self-selection: evidence from the German reunification experiment," *The Quarterly Journal of Economics*, 120(3), 1085–1120.

Ge, S., D. T. Yang, and J. Zhang (2012): "Population Policies, Demographic Structural Changes, and the Chinese Household Saving Puzzle," .

Gentry, W. M., and R. G. Hubbard (2000): "Tax policy and entrepreneurial entry," *American Economic Review*, 90(2), 283–287.

Giles, J., and K. Yoo (2007): "Precautionary behavior, migrant networks, and household consumption decisions: An empirical analysis using household panel data from rural China," *The Review of Economics and Statistics*, 89(3), 534–551.

Grawe, N. D. (2006): "Lifecycle bias in estimates of intergenerational earnings persistence," *Labour economics*, 13(5), 551–570.

Haider, S., and G. Solon (2006): "Life-cycle variation in the association between current and lifetime earnings," *American Economic Review*, 96(4), 1308–1320.

Hannum, E., and W. Meiyan (2006): "Geography and educational inequality in China," *China Economic Review*, 17(3), 253–265.

He, H., F. Huang, Z. Liu, and D. Zhu (2017): "Breaking the iron rice bowl: Evidence of precautionary savings from the chinese state-owned enterprises reform," *Journal of Monetary Economics*.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161–1189.

Horioka, C. Y., and A. Terada-Hagiwara (2017): "The impact of sex ratios before marriage on household saving in two Asian countries: The competitive saving motive revisited," *Review of Economics of the Household*, 15(3), 739–757.

Horioka, C. Y., and J. Wan (2007): "The determinants of household saving in China: a dynamic panel analysis of provincial data," *Journal of Money, Credit and Banking*, 39(8), 2077–2096.

Hurst, E., and A. Lusardi (2004): "Liquidity constraints, household wealth, and entrepreneurship," *Journal of political Economy*, 112(2), 319–347.

İmrohoroğlu, A., K. Zhao, et al. (2017): "The Chinese Saving Rate: Long-Term Care Risks, Family Insurance, and Demographics," Discussion paper.

Jia, R., and H. Li (2016): "Access to Elite Education, Wage Premium, and Social Mobility: The Truth and Illusion of China's College Entrance Exam," .

Keynes, J. M. (2016): *General theory of employment, interest and money*. Atlantic Publishers & Dist.

Kimball, M. S. (1990): "Precautionary Saving in the Small and in the Large," *Econometrica: Journal of the Econometric Society*, pp. 53–73.

Knight, J., S. Li, and Q. Deng (2010): "Son preference and household income in rural China," *The journal of development studies*, 46(10), 1786–1805.

Li, H., P. W. Liu, and J. Zhang (2012): "Estimating returns to education using twins in urban China," *Journal of Development Economics*, 97(2), 494–504.

Li, H., P. Loyalka, S. Rozelle, B. Wu, and J. Xie (2015): "Unequal access to college in China: How far have poor, rural students been left behind?," *The China Quarterly*, 221, 185–207.

Li, W., and J. Yi (2015): "The Competitive Earning Incentive for Sons: Evidence from Migration in China," Discussion paper, IZA Discussion Paper.

Liu, C., L. Zhang, R. Luo, X. Wang, S. Rozelle, B. Sharbono, J. Adams, Y. Shi, A. Yue, H. Li, et al. (2011): "Early commitment on financial aid and college decision making of poor students: Evidence from a randomized evaluation in rural China," *Economics of Education Review*, 30(4), 627–640.

Long, M. C. (2004): "Race and college admissions: An alternative to affirmative action?," *The Review of Economics and Statistics*, 86(4), 1020–1033.

Loyalka, P., J. Chu, J. Wei, N. Johnson, and J. Reniker (2017): "Inequalities in the pathway to college in China: when do students from poor areas fall behind?," *The China Quarterly*, 229, 172–194.

Loyalka, P., C. Liu, Y. Song, H. Yi, X. Huang, J. Wei, L. Zhang, Y. Shi, J. Chu, and S. Rozelle (2013): "Can information and counseling help students from poor rural areas go to high school? Evidence from China," *Journal of Comparative Economics*, 41(4), 1012–1025.

Loyalka, P., J. Wei, Y. Song, W. Zhong, and J. Chu (2014): "The impacts of building elite high schools for students from disadvantaged areas," *Economic Development and Cultural Change*, 63(2), 393–422.

Lusardi, A. (1996): "Permanent income, current income, and consumption: Evidence from two panel data sets," *Journal of Business & Economic Statistics*, 14(1), 81–90.

Meng, X. (2003): "Unemployment, consumption smoothing, and precautionary saving in urban China," *Journal of Comparative Economics*, 31(3), 465–485.

Meng, X., K. Shen, and S. Xue (2013): "Economic reform, education expansion, and earnings inequality for urban males in China, 1988–2009," *Journal of Comparative Economics*, 41(1), 227–244.

Mo, D., L. Zhang, H. Yi, R. Luo, S. Rozelle, and C. Brinton (2013): "School dropouts and conditional cash transfers: Evidence from a randomised controlled trial in rural China's junior high schools," *Journal of Development Studies*, 49(2), 190–207.

Modigliani, F., and S. L. Cao (2004): "The Chinese saving puzzle and the life-cycle hypothesis," *Journal of economic literature*, 42(1), 145–170.

Paxson, C. H. (1992): "Using weather variability to estimate the response of savings to transitory income in Thailand," *The American Economic Review*, pp. 15–33.

Quadrini, V. (1999): "The importance of entrepreneurship for wealth concentration and mobility," *Review of income and Wealth*, 45(1), 1–19.

——— (2000): "Entrepreneurship, saving, and social mobility," *Review of Economic Dynamics*, 3(1), 1–40.

Rosenbaum, P. R., and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

Rosenzweig, M., and J. Zhang (2014): "Co-residence, life-cycle savings and intergenerational support in urban China," Discussion paper, National Bureau of Economic Research.

Sautman, B. (1998): "Affirmative action, ethnic minorities and China's universities," *Pacific Rim Law Policy Journal*, 7, 77.

Schmidt-Hebbel, K., and L. Serven (2000): "Does income inequality raise aggregate saving?," *Journal of Development Economics*, 61(2), 417–446.

Song, Z., and D. T. Yang (2012): "Life cycle earnings and saving in a fast-growing economy," *Unpublished manuscript, Booth Sch. Bus., Univ. Chicago*.

Souleles, N. S. (1999): "The response of household consumption to income tax refunds," *American Economic Review*, 89(4), 947–958.

Stiglitz, J. E., and A. Weiss (1981): "Credit rationing in markets with imperfect information," *The American economic review*, 71(3), 393–410.

Tan, Z., S.-J. Wei, and X. Zhang (2015): "Deadly Discrimination: Implications of "Missing Girls" for Workplace Safety," Discussion paper, IFPRI Conference Paper.

Wang, X., C. Liu, L. Zhang, R. Luo, T. Glauben, Y. Shi, S. Rozelle, and B. Sharbono (2011): "College education and the poor in China: documenting the hurdles to educational attainment and college matriculation," *Asia Pacific Education Review*, 12(4), 533.

Wang, X., and Y. Wen (2012): "Housing prices and the high Chinese saving rate puzzle," *China Economic Review*, 23(2), 265–283.

Wei, S.-J., and X. Zhang (2011a): "The competitive saving motive: Evidence from rising sex ratios and savings rates in China," *Journal of political Economy*, 119(3), 511–564.

——— (2011b): "Sex ratios, entrepreneurship, and economic growth in the People's Republic of China," Discussion paper, National Bureau of Economic Research.

——— (2016): "The competitive saving motive: Concept, evidence, and implications," *Frontiers of Economics in China*, 11(3), 355–366.

Wei, S.-J., X. Zhang, and Y. Liu (2017): "Home ownership as status competition: Some theory and evidence," *Journal of Development Economics*, 127, 169–186.

Wu, X., and Z. Zhang (2010): "Changes in educational inequality in China, 1990–2005: Evidence from the population census data," in *Globalization, changing demographics, and educational challenges in East Asia*, pp. 123–152. Emerald Group Publishing Limited.

Yang, D. T., J. Zhang, and S. Zhou (2012): "Why are saving rates so high in China?," in *Capitalizing China*, pp. 249–278. University of Chicago Press.

Ye, H. (2015): "Key-point schools and entry into tertiary education in China," *Chinese Sociological Review*, 47(2), 128–153.

Zeldes, S. P. (1989a): "Consumption and liquidity constraints: an empirical investigation," *The Journal of Political Economy*, pp. 305–346.

——— (1989b): "Optimal consumption with stochastic income: Deviations from certainty equivalence," *The Quarterly Journal of Economics*, 104(2), 275–298.

Zhang, J., Y. Zhao, A. Park, and X. Song (2005): "Economic returns to schooling in urban China, 1988 to 2001," *Journal of Comparative Economics*, 33(4), 730–752.

Zhang, X., and R. Kanbur (2005): "Spatial inequality in education and health care in China," *China Economic Review*, 16(2), 189–204.

Zhou, M., and A. Hill (2009): *Affirmative action in China and the US: A dialogue on inequality and minority education*. Springer.

Zhou, W. (2014): "Brothers, household financial markets and savings rate in China," *Journal of Development Economics*, 111, 34–47.

Zhou, X., P. Moen, and N. B. Tuma (1998): "Educational stratification in urban China:

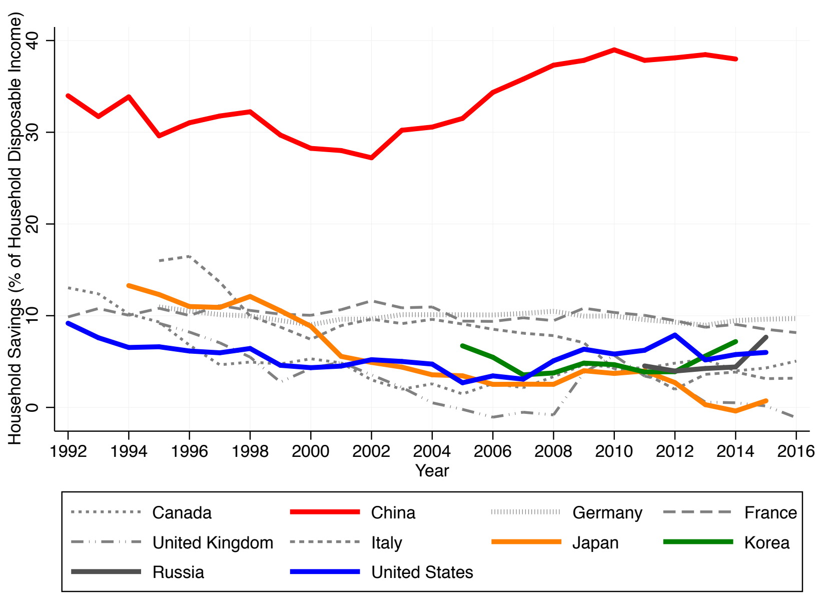
1949-94," *Sociology of Education*, pp. 199–222.

APPENDIX A

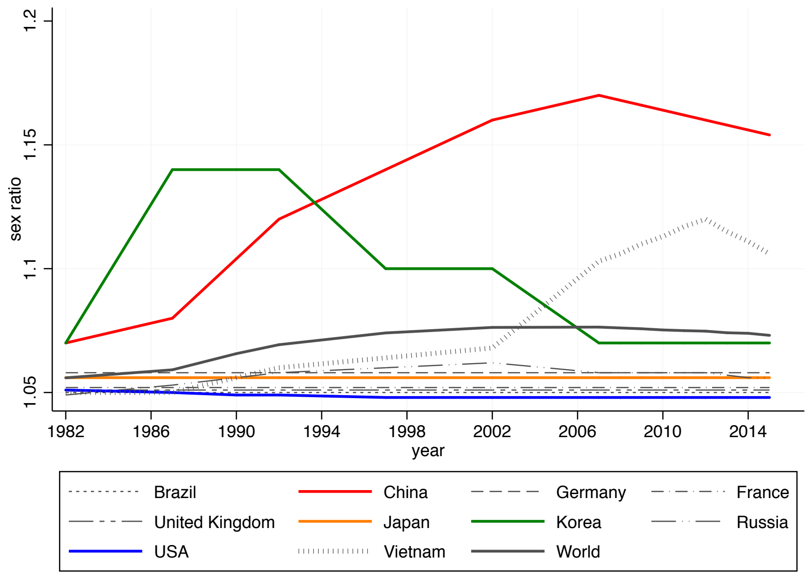
COMPETITIVE SAVING HYPOTHESIS REVISITED

A.1 Figures

Figure A.1: Aggregate saving rate and sex ratio at birth across countries



(1) Saving rate comparison across countries



(2) Sex ratio comparison across countries

Figure A.2: China's age 10-19 sex ratio and household saving rate

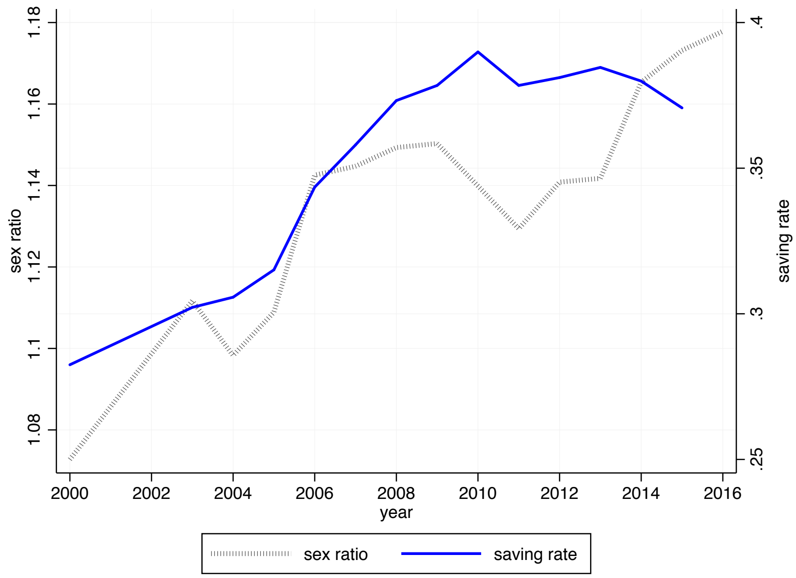


Figure A.3: Change in the distribution of sex ratio between 1990 and 2000: distribution comparison

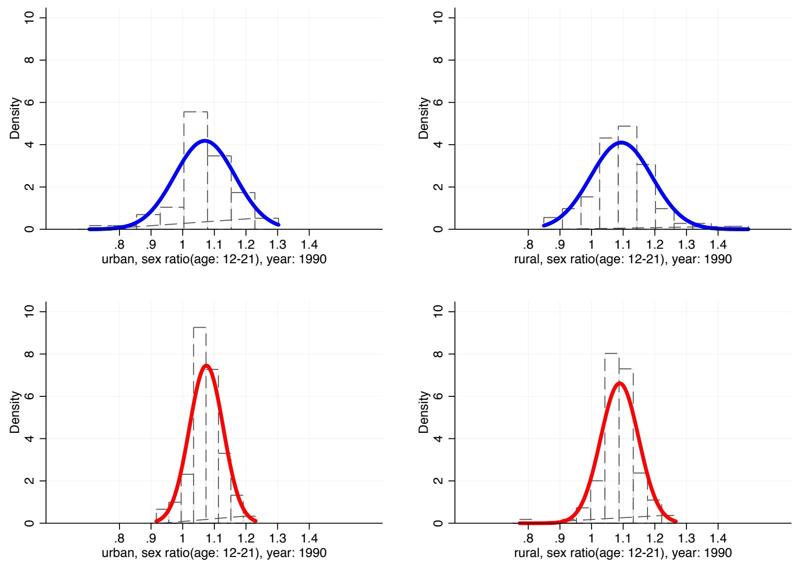
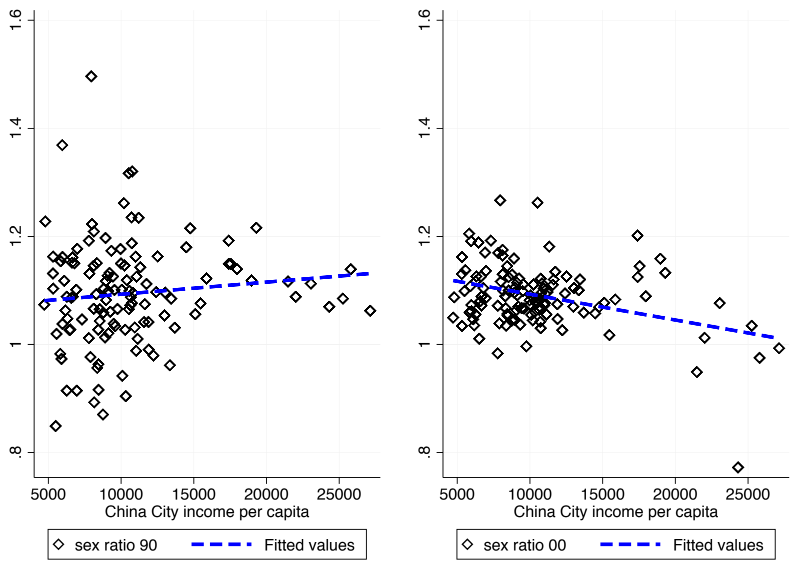
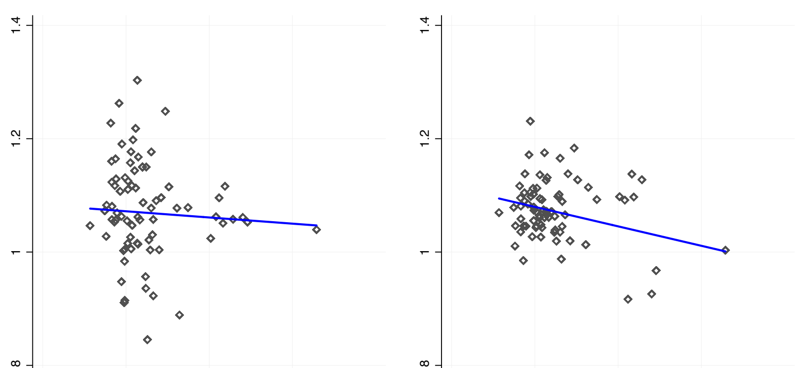


Figure A.4: Sex ratio vs. per capita income 1990 and 2000



(1) rural



(2) urban

Figure A.5: Change in the rural county sex ratio vs. per capita GDP and income

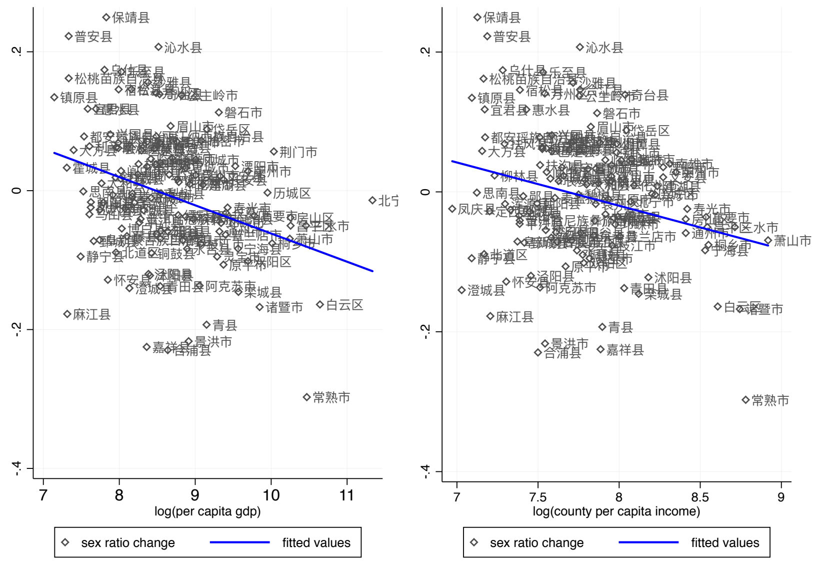


Figure A.6: Saving rate and sex ratio by households in rich and poor counties between 1990 and 2000

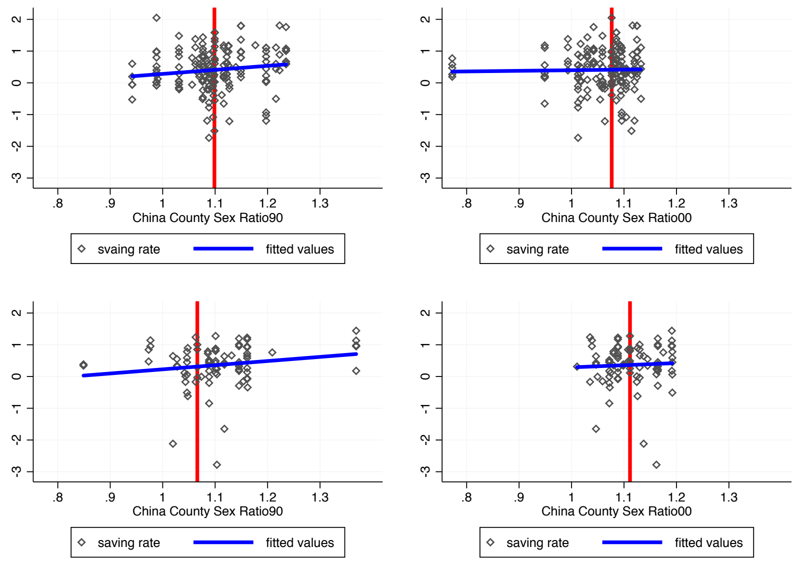
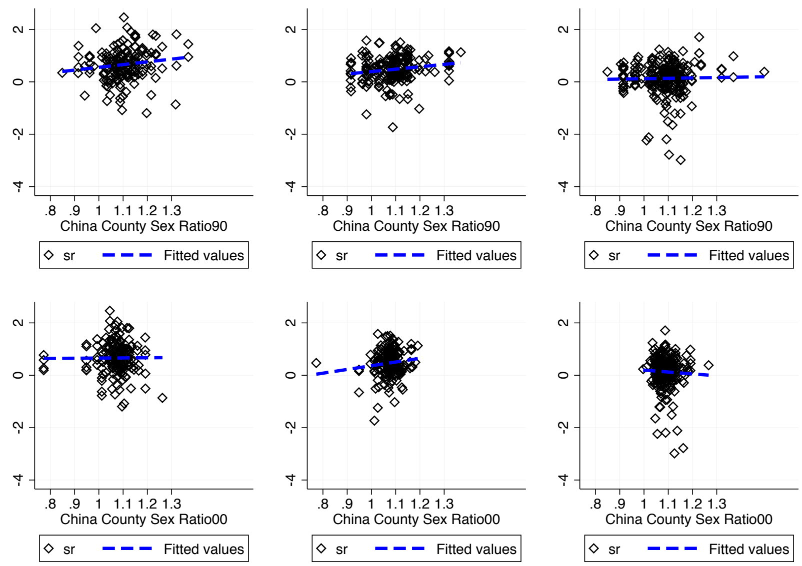
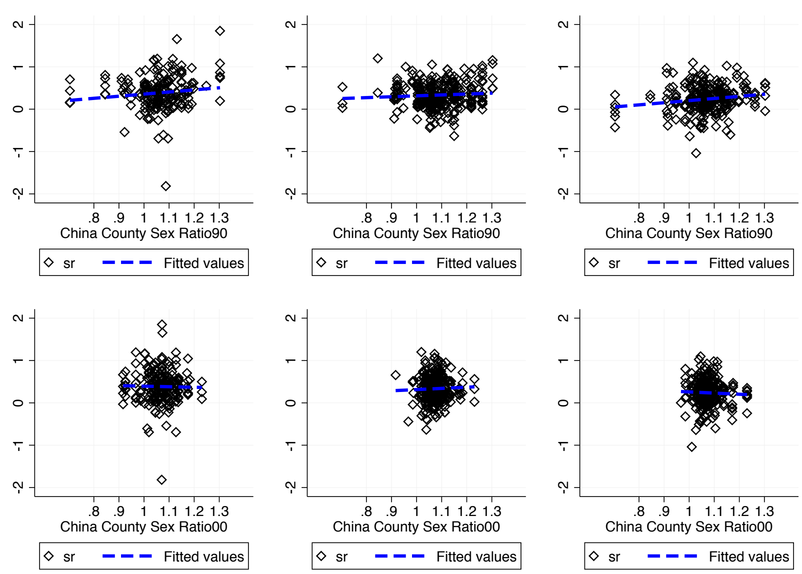


Figure A.7: Saving rate and sex ratio by rich, middle-income, and poor households rich between 1990 and 2000



(1) rural



(2) urban

A.2 Tables

Table A.1: Summary statistics: a comparison with Wei and Zhang (2011)

Household type	Mean	Median	Max	Min	SD	Observations
Rural (Wei and Zhang, 2011)						
One son	0.393	0.394	2.462	-2.986	0.625	580
One daughter	0.318	0.353	1.812	-3.559	0.626	326
sex ratio 1990	1.090		1.230	1.010	0.040	122
Urban (Wei and Zhang, 2011)						
One son	0.312	0.306	1.849	-1.816	0.333	769
One daughter	0.302	0.308	2.153	-1.299	0.356	766
sex ratio 1990	1.080		1.240	1.020	0.040	77
Rural (this paper)						
One son	0.387	0.392	2.462	-2.986	0.624	587
One daughter	0.325	0.365	1.821	-3.559	0.622	329
sex ratio 1990	1.090	1.094	1.490	0.850	0.970	122
Urban (this paper)						
One son	0.314	0.31	1.849	-1.816	0.332	764
One daughter	0.305	0.311	2.153	-1.299	0.357	763
sex ratio 1990	1.070	1.063	1.300	0.700	0.090	77

Table A.2: Results comparison with Wei and Zhang (2011): local sex ratios for the age cohort 12-21 in 2002 from 1990 census 0-9 years old

(a) Rural household-level evidence

	Subsample that removes the following potential outliers: Rural							
	Full sample		inc. or expend. < 2,000		Bottom & Top 5%		Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
sex ratio	0.54**	-0.13	0.52**	-0.27	0.62***	-0.24	0.65***	-0.21
(county)	(0.23)	(0.27)	(0.23)	(0.23)	(0.17)	(0.19)	(0.20)	(0.26)
Obersvations	586	329	569	318	528	297	446	242
sex ratio	1.34**	-0.17	1.38**	-0.18	1.10***	-0.23	1.20***	-0.32
(county)	(0.52)	(0.55)	(0.51)	(0.54)	(0.44)	(0.43)	(0.43)	(0.44)
Observations	580	326	564	315	522	292	489	269

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(b) Urban household-level evidence

	Subsample that removes the following potential outliers: Urban							
	Full sample		inc. or expend. < 3,000		Bottom & Top 5%		Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
sex ratio	0.38***	0.24*	0.36***	0.22*	0.17**	0.13*	0.21**	0.10
(county)	(0.12)	(0.13)	(0.12)	(0.13)	(0.07)	(0.08)	(0.09)	(0.13)
Observations	764	763	762	759	688	687	377	396
sex ratio	1.54**	1.85**	1.16**	1.07**	0.74***	0.65**	0.98**	0.47
(county)	(0.29)	(0.33)	(0.30)	(0.37)	(0.24)	(0.26)	(0.31)	(0.44)
Obersvations	769	766	604	605	691	688	384	399

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Robustness checks using other county characteristics

(a) Per capita GDP may have no effect on the competitive motive

	Subsample that removes the following potential outliers: Urban							
	Full sample		inc. or expend. < 2,000		Bottom & Top 5%		Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
sex ratio 1990	0.53**	-0.09	0.51**	-0.24	0.61***	-0.21	0.65***	-0.20
(county)	(0.23)	(0.26)	(0.22)	(0.23)	(0.17)	(0.19)	(0.19)	(0.26)
ln_pcgdp	-0.08**	-0.12***	-0.07**	-0.11***	-0.06**	-0.07**	-0.04	-0.07*
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)
Observations	586	329	569	318	528	297	446	242
sex ratio 1990	0.36***	0.23*	0.34***	0.21*	0.16*	0.12*	0.26**	0.11
(county)	(0.12)	(0.13)	(0.12)	(0.13)	(0.09)	(0.07)	(0.13)	(0.13)
ln_pcgdp	-0.04***	-0.06***	-0.05***	-0.06***	-0.02	-0.04***	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Observations	764	763	762	759	688	687	377	396

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(b) Urbanization may have no effect on the competitive motive

	Subsample that removes the following potential outliers: Urban							
	Full sample		inc. or expend. < 2,000		Bottom & Top 5%		Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
ratio_1221_90	0.49**	-0.13	0.48**	-0.28	0.58***	-0.23	0.62***	-0.22
	(0.23)	(0.27)	(0.22)	(0.22)	(0.17)	(0.19)	(0.19)	(0.26)
urban_ratio	-0.47**	-0.45**	-0.45**	-0.51**	-0.36**	-0.34**	-0.30**	-0.30
	(0.19)	(0.21)	(0.18)	(0.20)	(0.14)	(0.17)	(0.14)	(0.19)
Observations	586	329	569	318	528	297	446	242
ratio_1221_90	0.38***	0.23*	0.36***	0.21*	0.16*	0.04	0.25**	0.12
	(0.12)	(0.13)	(0.12)	(0.13)	(0.09)	(0.10)	(0.13)	(0.13)
urban_ratio	-0.30***	-0.27***	-0.30***	-0.27***	-0.16***	-0.15***	-0.10*	-0.12*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.07)
Observations	764	763	762	759	688	687	377	396

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(c) Provincial migration share may have no effect on the competitive motive

	Subsample that removes the following potential outliers:						Urban	
	Full sample		inc. or expend. < 2,000		Bottom & Top 5%		Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
ratio_1221_90	0.53**	-0.12	0.52**	-0.26	0.62***	-0.21	0.64***	-0.20
	(0.23)	(0.27)	(0.22)	(0.23)	(0.17)	(0.19)	(0.20)	(0.26)
migration_prov	-4.02***	-2.06***	-3.93***	-2.21***	-2.23***	-1.60***	-1.76**	-1.59***
	(0.96)	(0.58)	(0.94)	(0.57)	(0.69)	(0.45)	(0.69)	(0.49)
Observations	586	329	569	318	528	297	446	242
ratio_1221_90	0.37***	0.23*	0.36***	0.21	0.16*	0.04	0.24*	0.10
	(0.12)	(0.13)	(0.12)	(0.13)	(0.09)	(0.10)	(0.13)	(0.13)
migration_prov	-0.64**	-0.58**	-0.70**	-0.62**	-0.46**	-0.25	0.04	-0.19
	(0.29)	(0.29)	(0.28)	(0.29)	(0.20)	(0.21)	(0.22)	(0.24)
Observations	764	763	762	759	688	687	377	396

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(d) City migration share have no effect on the competitive motive

	Subsample that removes the following potential outliers:						Urban	
	Full sample		inc. or expend. < 2,000		Bottom & Top 5%		Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
ratio_1221_90	0.56**	-0.11	0.54**	-0.25	0.63***	-0.21	0.65***	-0.20
	(0.23)	(0.27)	(0.22)	(0.23)	(0.17)	(0.19)	(0.19)	(0.26)
migration_city	-3.06***	-1.65***	-2.96***	-1.77***	-1.87***	-1.24***	-1.55***	-1.25***
	(0.73)	(0.49)	(0.71)	(0.48)	(0.53)	(0.34)	(0.52)	(0.36)
Observations	586	329	569	318	528	297	446	242
ratio_1221_90	0.37***	0.22*	0.36***	0.21	0.16*	0.04	0.24*	0.10
	(0.12)	(0.13)	(0.12)	(0.13)	(0.09)	(0.10)	(0.13)	(0.13)
migration_city	-0.52***	-0.42**	-0.56***	-0.44**	-0.34**	-0.21	-0.00	-0.11
	(0.18)	(0.20)	(0.17)	(0.20)	(0.13)	(0.14)	(0.15)	(0.17)
Observations	764	763	762	759	688	687	377	396

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(e) Unmarried population share may have no effect on the competitive motive

	Full sample		Subsample that removes the following potential outliers: inc. or expend. < 2,000		Bottom & Top 5%		Urban Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
ratio_1221_90	0.55**	-0.01	0.53**	-0.18	0.59***	-0.23	0.63***	-0.19
	(0.24)	(0.29)	(0.23)	(0.24)	(0.18)	(0.21)	(0.20)	(0.27)
unmarried_15	-0.14	-1.17*	-0.06	-0.81	0.36	-0.08	0.31	-0.24
	(0.67)	(0.60)	(0.64)	(0.56)	(0.49)	(0.54)	(0.53)	(0.58)
Observations	586	329	569	318	528	297	446	242
ratio_1221_90	0.34***	0.19	0.32***	0.17	0.17*	0.01	0.25**	0.10
	(0.12)	(0.14)	(0.12)	(0.13)	(0.10)	(0.10)	(0.13)	(0.13)
unmarried_15	0.45	0.45	0.55	0.36	-0.05	0.31	-0.52	-0.09
	(0.36)	(0.36)	(0.36)	(0.36)	(0.27)	(0.29)	(0.38)	(0.40)
Observations	764	763	762	759	688	687	377	396

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: What if use year 2000 sex ratio rather than 1990? same specification as Wei and Zhang (2011)

	Full sample		Subsample that removes the following potential outliers: inc. or expend. < 3,000 (< 5,000 for urban)		Bottom & Top 5%		Rural & Urban Bottom & Top 5% no explicit marriage status	
	son	daughter	son	daughter	son	daughter	son	daughter
sex ratio 2000	0.19	0.38	0.42	0.17	0.23	-0.20	0.18	-0.09
(county level)	(0.45)	(0.46)	(0.43)	(0.44)	(0.31)	(0.38)	(0.33)	(0.43)
Observations	586	329	569	318	528	297	446	242
sex ratio 2000	0.17	-0.02	0.15	-0.00	-0.04	-0.37**	-0.02	-0.59**
(county level)	(0.20)	(0.23)	(0.20)	(0.23)	(0.16)	(0.18)	(0.21)	(0.24)
Observations	764	763	762	759	688	687	377	396

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Summary statistics of sex ratio 2000 vs. sex ratio 1990

Rural Counties	Mean	Max	Min	SD	Observations
ratio age0-9, census 1990	1.09	1.49	0.85	0.09	122
ratio age0-9, census 1990 (Wei and Zhang, 2011)	1.09	1.23	1.01	0.04	122
ratio age10-19, census 2000	1.09	1.27	0.77	0.06	122
Urban Cities	**Mean**	**Max**	**Min**	**SD**	**Observations**
ratio age0-9, census 1990	1.07	1.30	0.70	0.09	77
ratio age0-9, census 1990 (Wei and Zhang, 2011)	1.08	1.24	1.02	0.04	77
ratio age10-19, census 2000	1.07	1.23	0.92	0.05	77

Table A.6: Migration for the specific age12-21 group: the rich rural vs. the poor rural

	Full sample		Subsample that removes outliers	
	son	daughter	sr outliers of son	sr outliers of daughter
sex ratio 0-9	1.11*	0.01	1.13*	-0.18
(county 1990)	(0.64)	(0.56)	(0.64)	(0.55)
sex ratio 10-19	0.90*	0.50	0.87*	0.46
(county 2000)	(0.52)	(0.57)	(0.53)	(0.59)
sex ratio 0-9	0.93*	-1.13	1.49**	-0.84
(county 1990)	(0.80)	(0.77)	(0.71)	(0.69)
sex ratio 10-19	-2.34*	-0.69	-1.26	-1.02
(county 2000)	(1.31)	(1.45)	(1.31)	(2.07)

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Results comparisons by county and household income

(a) rural

rural	data sample variables	sex ratio 1990 census son s.r.	sex ratio 1990 census daughter s.r.	sex ratio 2000 census son s.r.	sex ratio 2000 census daughter s.r.
full sample	ratio_1221_90/00	0.54**	-0.13	0.19	0.38
		(0.23)	(0.27)	(0.45)	(0.46)
rich county	ratio_1221_90/00	1.08**	0.80	1.15***	0.53
		(0.53)	(0.66)	(0.43)	(0.54)
middle county	ratio_1221_90/00	0.44	-0.58	-2.22**	-1.22
		(0.33)	(0.41)	(0.95)	(0.98)
poor county	ratio_1221_90/00	0.56	0.01	-2.29	0.03
		(0.54)	(0.50)	(1.49)	(1.38)
rich household	ratio_1221_90/00	1.00*	1.48	-0.07	0.46
		(0.56)	(1.01)	(0.64)	(0.60)
middle household	ratio_1221_90/00	0.90**	-0.07	1.44*	-1.68
		(0.35)	(0.45)	(0.83)	(1.97)
poor household	ratio_1221_90/00	0.00	-0.43	-1.48	0.69
		(0.33)	(0.35)	(0.93)	(0.80)

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(b) urban

urban	data sample variables	sex ratio 1990 census son s.r.	sex ratio 1990 census daughter s.r.	sex ratio 2000 census son s.r.	sex ratio 2000 census daughter s.r.
full sample	ratio_1221_90/00	0.38***	0.24*	0.17	-0.02
		(0.12)	(0.13)	(0.20)	(0.23)
rich county	ratio_1221_90/00	-0.23	0.10	0.68*	0.26
		(0.26)	(0.25)	(0.36)	(0.38)
middle county	ratio_1221_90/00	0.53***	-0.03	0.56	-0.08
		(0.18)	(0.15)	(0.48)	(0.45)
poor county	ratio_1221_90/00	0.64***	1.00***	-0.60	-0.86**
		(0.24)	(0.28)	(0.39)	(0.41)
rich household	ratio_1221_90/00	0.57*	0.30	0.45	0.40
		(0.30)	(0.26)	(0.30)	(0.39)
middle household	ratio_1221_90/00	0.16	0.23	0.22	0.13
		(0.18)	(0.22)	(0.43)	(0.48)
poor household	ratio_1221_90/00	0.48**	0.23	-0.01	-0.62*
		(0.19)	(0.19)	(0.38)	(0.36)

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Marriage migration: major form of migration and cause of female migration

Marriage as a Cause of Migration (1990)(%)			
Destination	Both sexes	Men	Women
Intra-provincial			
All	14.0	2.0	28.0
City	3.0	1.0	6.0
Town	8.0	1.0	15.0
Rural areas	19.0	3.0	35.0
Inter-provincial			
All	14.0	2.0	30.0
City	2.0	1.0	6.0
Town	8.0	1.0	18.0
Rural areas	20.0	3.0	40.0

Table A.9: Competitive hypothesis: evidence from CHFS 2015

	Full sample		Subsample that removes outliers	
	son	daughter	sr outliers of son	sr outliers of daughter
sex ratio 5-14	0.66**	0.75**	0.68**	0.66*
(county, 2010)	(0.31)	(0.37)	(0.32)	(0.37)
sex ratio 10-19	0.54	0.78*	0.61*	0.77*
(county, 2010)	(0.36)	(0.43)	(0.37)	(0.43)
Observations	796	635	752	604
sex ratio 5-14	0.04	0.22	0.03	0.19
(county, 2010)	(0.32)	(0.24)	(0.30)	(0.23)
sex ratio 10-19	0.04	0.35	0.04	0.37
(county, 2010)	(0.23)	(0.26)	(0.23)	(0.26)
Observations	2,010	1,556	1,915	1,484

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Sex ratios and saving rates across province: dependent variable=(Y-C)/Y

	full	urban	rural	full	urban	rural
sex ratio for age cohort 5-19	-0.03	0.05	0.24*			
	(0.10)	(0.09)	(0.14)			
sex ratio for age cohort 10-19				-0.03	0.04	0.21*
				(0.08)	(0.07)	(0.12)
Per capita income (log)	0.20***	0.14***	0.39***	0.20***	0.15***	0.39***
	(0.03)	(0.03)	(0.05)	(0.03)	(0.03)	(0.05)
share of population of aged 0-14	-0.25	0.04	0.80	-0.25	0.04	0.81
	(0.49)	(0.42)	(0.70)	(0.49)	(0.42)	(0.70)
share of population of aged 15-64	-0.37	0.03	0.22	-0.37	0.02	0.21
	(0.46)	(0.39)	(0.65)	(0.46)	(0.39)	(0.65)
share of SOE emplyoment in total labor force	-0.05	-0.21***	0.04	-0.05	-0.21***	0.03
	(0.07)	(0.06)	(0.10)	(0.07)	(0.06)	(0.10)
Constant	-0.74	-0.87*	-2.76***	-0.74	-0.86*	-2.74***
	(0.55)	(0.49)	(0.77)	(0.55)	(0.49)	(0.77)
provincial fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	214	214	214	214	214	214
R^2	0.77	0.85	0.74	0.77	0.85	0.74

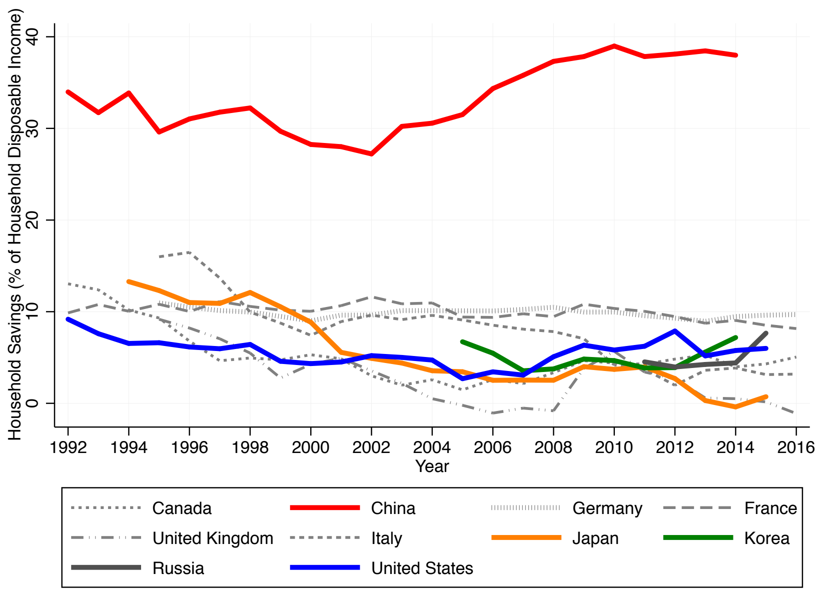
Notes: 1. Robust standard errors (cluster at provincial level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

APPENDIX B

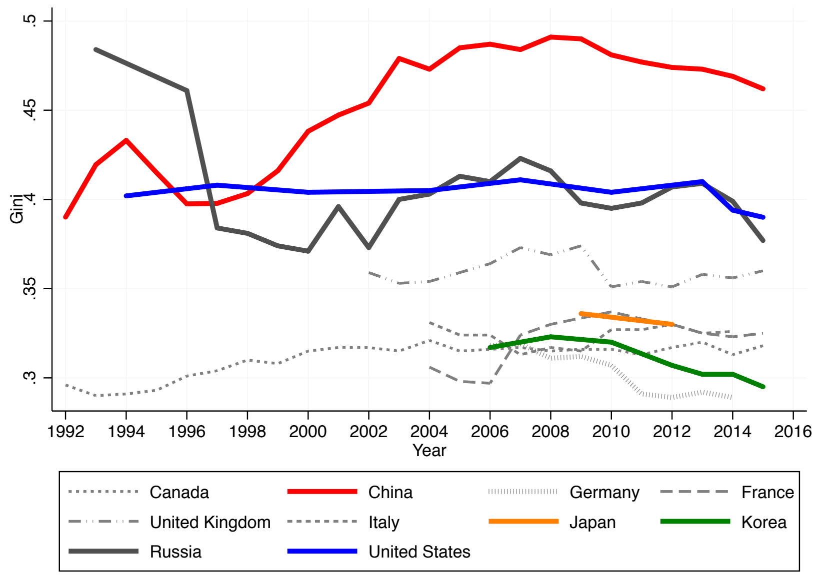
INCOME INEQUALITY, LIQUIDITY CONSTRAINTS, AND CHINA'S HOUSEHOLD SAVING RATE

B.1 Figures

Figure B.1: International comparison

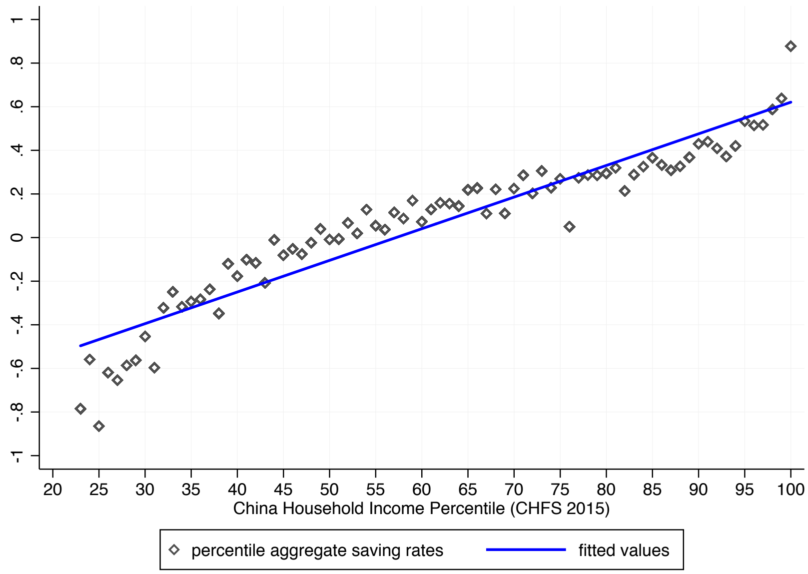


(1) Saving rate comparison across countries

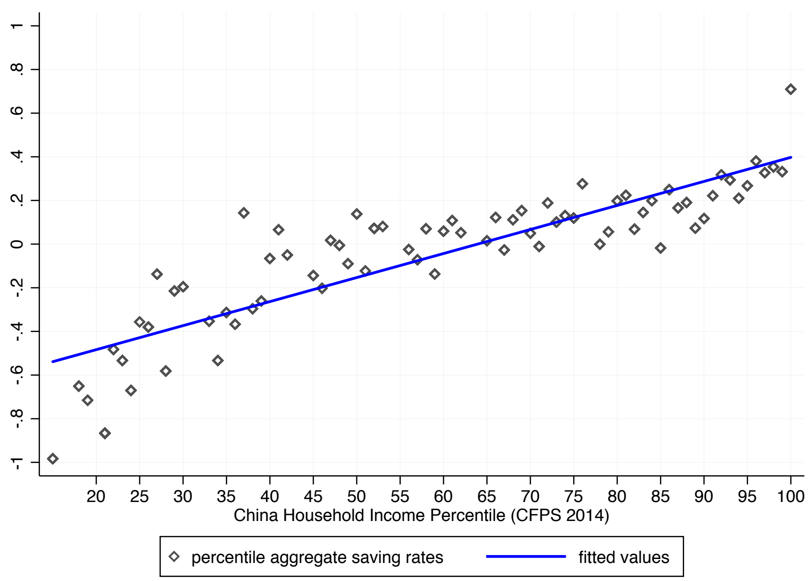


(2) Gini coefficient across countries

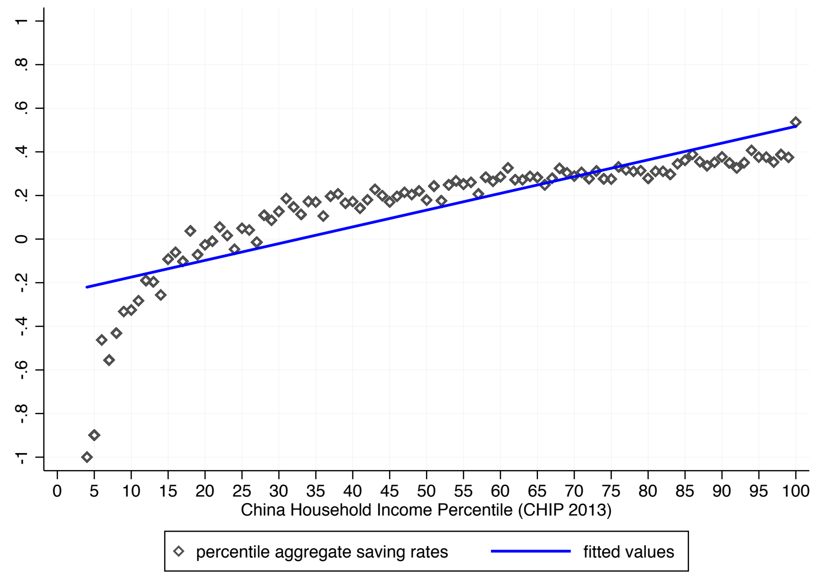
Figure B.2: Chinese household saving rate by income class



(1) CHFS 2015



(2) CFPS 2014



1
.8
.6
.4
.2
0
-.2
-.4
-.6
-.8
-1
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
China Household Income Percentile (CHIP 2013)
percentile aggregate saving rates
fitted values

(3) CHIP 2013

Figure B.3: China's household saving rate and Gini coefficient

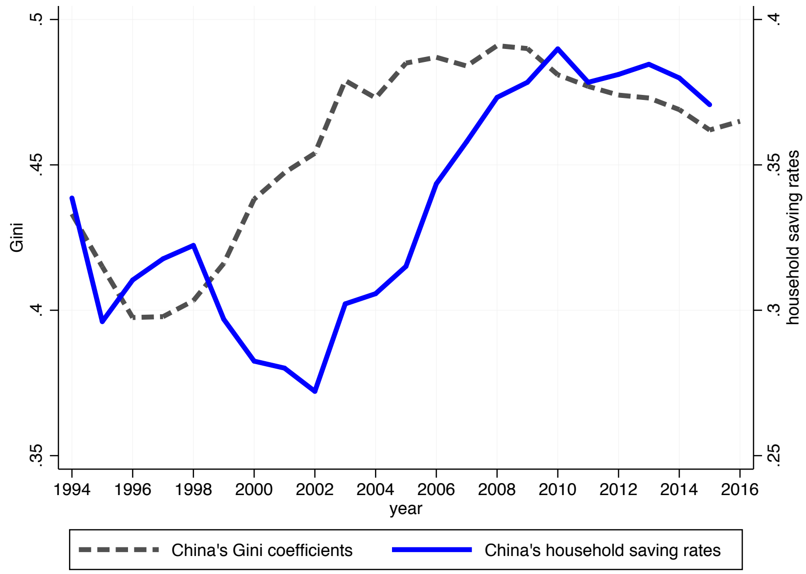
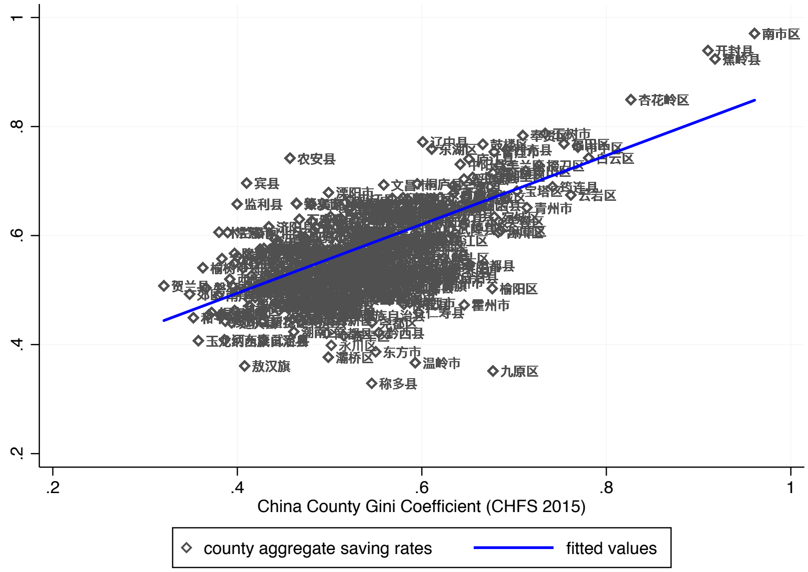
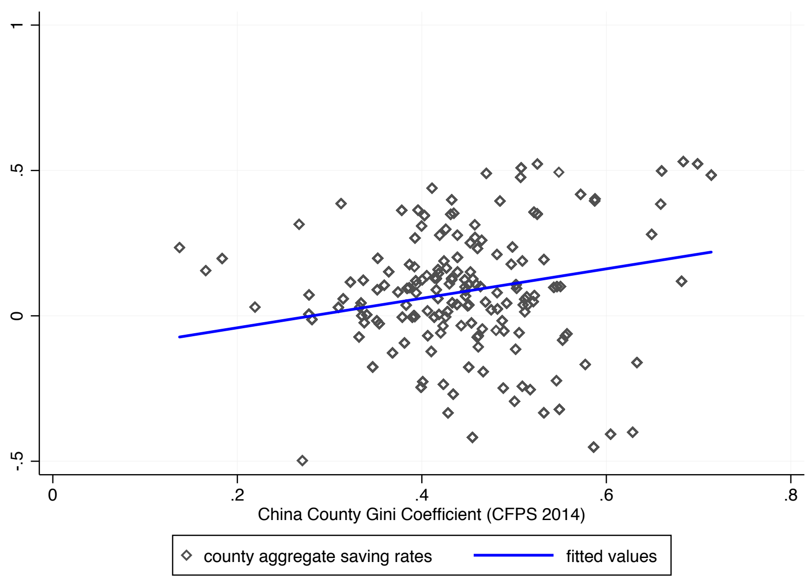


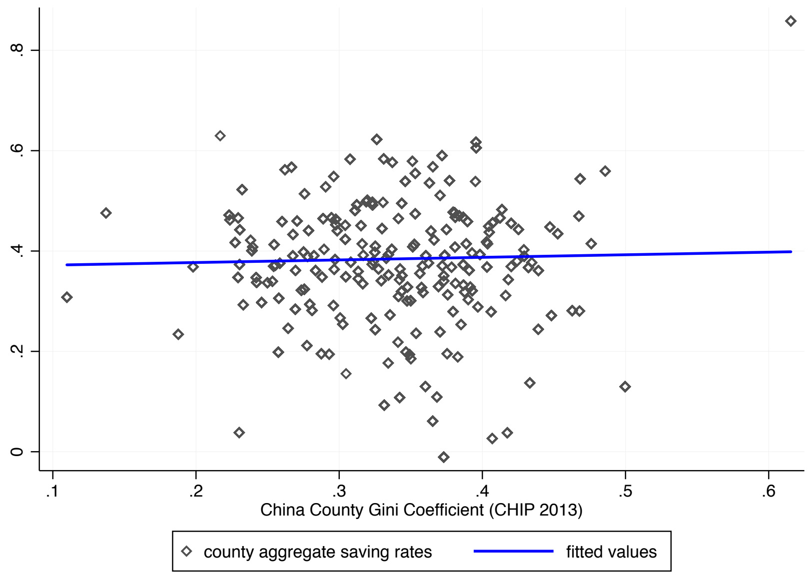
Figure B.4: County-level aggregate saving rates and Gini coefficients



(1) CHFS 2015

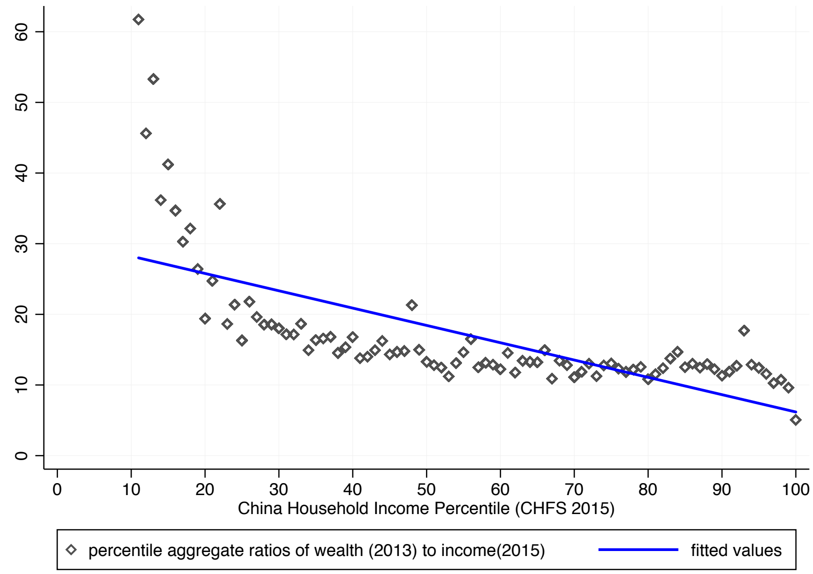


(2) CFPS 2014

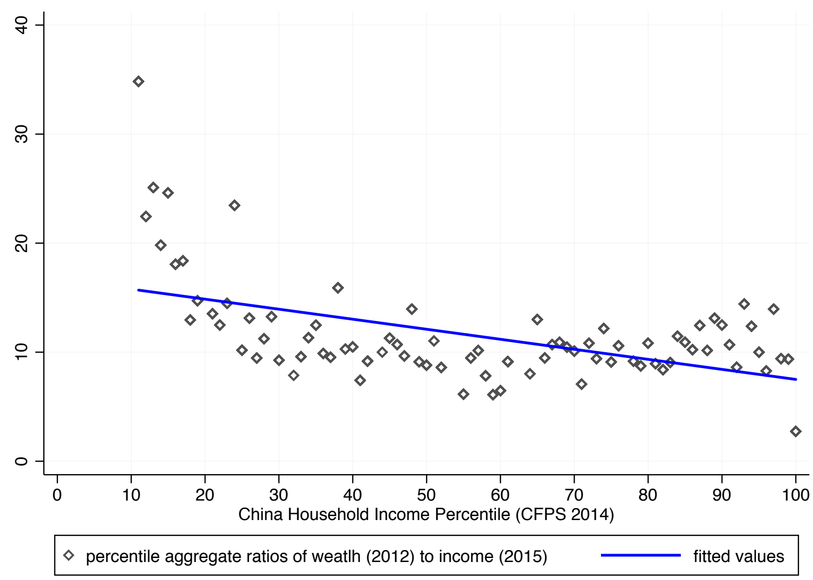

.8
.6
.4
.2
0
.1
.2
.3
.4
.5
.6
China County Gini Coefficient (CHIP 2013)
county aggregate saving rates
fitted values

(3) CHIP 2013

Figure B.5: The ratio of previous wealth to current income by income percentile



(1) CHFS



(2) CFPS

B.2 Tables

Table B.1: The uneven distribution of China's household saving rate

	CHFS 2015		CFPS 2014		CHIP 2013		NBS
	saving rate	shares	saving rate	shares	saving rate	shares	saving rate
top 1%	0.866	69.1	0.581	44.0	0.536	12.2	
top 5%	0.741	99.7	0.515	76.7	0.426	29.3	
top 10%	0.672	116.0	0.471	96.6	0.403	44.4	
top 25%	0.569	138.2	0.412	133.6	0.373	73.5	
bottom 50%	-1.327	-45.8	-0.634	-52.8	0.017	1.35	
% savers	55.9		45.3		74.6		
saving rate (2015)	0.291						0.285
saving rate (2014)			0.189				0.281
saving rate (2013)					0.273		0.278

Table B.2: Credit constraints and saving rates by income group

(a) CHFS 2015 Urban

	measure I			measure II		
	FA < 2 mon. PI		FA ≥ 2 mon. PI	no CC		at least one CC
	% constrained	s.r	s.r.	% constrained	s.r	s.r
top 20%	16.93	0.769	0.389	59.51	0.545	0.435
middle income	31.52	0.125	0.003	81.65	0.096	-0.179
bottom 20%	37.78	-0.794	-1.587	92.27	-1.142	-2.677

(b) CFPS 2014 Urban

	measure I		
	FA < 2 mon. PI		FA ≥ 2 mon. PI
	% constrained	s.r	s.r.
top 20%	43.01	0.338	0.242
middle income	43.04	-0.015	-0.071
bottom 20%	58.53	-0.781	-0.977

(c) CHIP 2013 Urban

	measure I		
	FA < 2 mon. PI		FA ≥ 2 mon. PI
	% constrained	s.r	s.r.
top 20%	15.81	0.422	0.403
middle income	14.40	0.388	0.306
bottom 20%	21.31	0.310	0.2595

Table B.3: Definition of variables and summary statistics

(a) Definition of variables

Variable	Description
household consumption	sum of family members' expenditure on food, clothing, housing, appliance and commodities, communication and transportation, culture recreation and entertainment, medical care and others
household income	sum of family members' wage income, business income, agricultural income, investment income and transfer income
hh_income_i	dummy variable, equals one if the household income is in the ith quintile and zero otherwise
hh_size	total number of members in the family
YDratio	number of age 0-14 in the household divided by hh_size
ODratio	number of age above 65 in the household divided by hh_size
employed	dummy variable, takes one if currently employed
employed_typ_j	dummy variable, equals one if household head works in type j organization, $j = 1$ (SOE, collective, or government), 2 (private or foreign company), or 3 (others)
hukou	dummy variable, takes one for urban and 0 for rural
hh_health	dummy variable, equals one if the household has a poor health member and zero otherwise
health_insurance	dummy variable, takes one if the household has health insurance and zero otherwise
pension	dummy variable, equals one if the household has pension and zero otherwise.
house_owner	dummy variable, takes one if the household own house and zero otherwise.
boy_number	number of boys in the household
girl_number	number of girls in the household
childage_04/59/1014/1519	dummy variables, takes one for child's age between 0 and 4 / between 5 and 9/ between 10 and 14/between 15 and 19)
age	age of household head
age2	age square
gender	dummy variable, takes one for male and zero for female
married	dummy variable, equals one for married household head and zero for unmarried household head
ccp_member	dummy variable, takes one for ccp member and zero otherwise
yos	years of schooling
hh_credit	dummy variable, equals one for constrained household and zero for unconstrained household

(b) Summary of statistics: CHFS 2015

Variable	Obs	Mean	Std.Dev.	Min	Max
hh size	25,635	3.177	1.449	1	20
YDratio	25,634	0.117	0.162	0	1
ODratio	25,634	0.170	0.318	0	1
employed	25,563	0.591	0.492	0	1
employ type	10,248	1.775	0.737	1	3
industry	9,342	11.28	6.968	1	24
hukou	24,377	0.607	0.488	0	1
health	25,584	0.186	0.389	0	1
hh health	25,628	0.314	0.464	0	1
health insurance	25,635	0.914	0.280	0	1
pension	25,169	0.792	0.406	0	1
house owner	22,802	0.990	0.0997	0	1
boy number	25,635	0.340	0.546	0	4
girl number	25,635	0.260	0.517	0	5
childage_04	25,635	0.134	0.377	0	7
childage_59	25,635	0.155	0.397	0	5
childage_1014	25,635	0.132	0.364	0	6
childage_1519	25,635	0.143	0.382	0	5
age	25,628	52.13	14.96	3	101
age2	25,628	2941	1607	9	10201
gender	25,635	0.699	0.459	0	1
married	23,304	0.834	0.372	0	1
ccp member	24,079	0.113	0.316	0	1
yos	25,598	10.28	4.098	0	22
hh credit	25,635	0.721	0.448	0	1

(c) Summary of statistics: CFPS 2014

Variable	Obs	Mean	Std.Dev.	Min	Max
hh size	6,603	3.005	1.578	1	14
YDratio	6,603	0.136	0.183	0	1
ODratio	6,603	0.150	0.306	0	1
employed	6,255	0.600	0.490	0	1
employ type	2,641	1.777	0.647	1	3
industry	3,117	8.100	5.188	1	19
hukou	6,387	0.514	0.500	0	1
health	6,597	0.170	0.376	0	1
hh health	6,598	0.273	0.446	0	1
health insurance	6,603	0.876	0.330	0	1
pension	6,603	0.406	0.491	0	1
house owner	6,599	0.821	0.383	0	1
boy number	6,603	0.359	0.573	0	4
girl number	6,603	0.303	0.566	0	5
childage_04	6,603	0.167	0.431	0	4
childage_59	6,603	0.166	0.430	0	4
childage_1014	6,603	0.143	0.379	0	3
childage_1519	6,603	0.154	0.396	0	3
age	6,603	54.37	16.30	0	102
age2	6,603	3221	1792	0	10404
gender	6,603	0.669	0.471	0	1
ethnicity	597	0.0704	0.256	0	1
married	6,600	0.767	0.423	0	1
ccp member	6,603	0.125	0.331	0	1
yos	6,600	7.987	4.870	0	19
schooling	6,600	1.441	0.708	1	3
hh credit	5,986	0.545	0.498	0	1

(d) Summary of statistics: CHIP 2013

Variable	Obs	Mean	Std.Dev.	Min	Max
hh size	6,674	2.980	1.107	1	8
YDratio	6,674	0.121	0.162	0	0.667
ODratio	6,674	0.121	0.283	0	1
employed	6,674	0.677	0.468	0	1
employ type	4,515	1.604	0.652	1	3
industry	4,513	9.779	5.993	1	20
hukou	6,672	0.844	0.363	0	1
health	6,667	0.0601	0.238	0	1
hh health	6,670	0.115	0.319	0	1
health insurance	6,640	1.733	6.005	0	156
pension	6,597	1.232	3.533	0	163
house owner	6,322	0.966	0.180	0	1
boy number	6,674	0.309	0.496	0	3
girl number	6,674	0.250	0.471	0	3
childage_04	6,674	0.109	0.324	0	2
childage_59	6,674	0.131	0.350	0	2
childage_1014	6,674	0.138	0.356	0	2
childage_1519	6,674	0.150	0.379	0	3
age	6,674	50.22	13.18	17	97
age2	6,674	2696	1400	289	9409
gender	6,674	0.728	0.445	0	1
ethnicity	6,673	0.0451	0.208	0	1
married	6,672	1.112	0.315	1	2
ccp member	6,634	0.276	0.447	0	1
yos	6,672	11.34	3.555	0	21
hh credit	6,331	0.840	0.366	0	1

Table B.4: The rich do save more

(a) Saving rate and current income

	CHFS 2015 Urban		CFPS 2014 Urban		CHIP 2013 Urban	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
hh_income_2	0.303***	0.305***	0.257***	0.247***	0.0731***	0.0868***
	(0.0157)	(0.0221)	(0.0203)	(0.0394)	(0.0139)	(0.0160)
hh_income_3	0.462***	0.453***	0.349***	0.292***	0.135***	0.166***
	(0.0149)	(0.0213)	(0.0215)	(0.0396)	(0.0143)	(0.0147)
hh_income_4	0.582***	0.553***	0.464***	0.385***	0.178***	0.186***
	(0.0154)	(0.0202)	(0.0224)	(0.0392)	(0.0142)	(0.0146)
hh_income_5	0.783***	0.750***	0.628***	0.544***	0.241***	0.262***
	(0.0159)	(0.0204)	(0.0238)	(0.0419)	(0.0131)	(0.0156)
hh_size	-0.0219***	-0.0214***	-0.0267***	-0.0498***	-0.000924	-0.00535
	(0.00459)	(0.00705)	(0.00844)	(0.0148)	(0.00604)	(0.00819)
YDratio	-0.0730	0.0245	-0.155*	-0.240*	-0.105**	-0.0499
	(0.0486)	(0.0609)	(0.0829)	(0.132)	(0.0462)	(0.0580)
ODratio	0.0656*	0.0989*	0.0653**	-0.115	0.0674***	0.0105
	(0.0395)	(0.0571)	(0.0264)	(0.115)	(0.0202)	(0.0484)
employed	0.0837***	0.0837***	0.00187	0.0397	0.0463***	0.0923
	(0.0108)	(0.0108)	(0.0154)	(0.0422)	(0.0138)	(0.107)
employ_typ_2		0.00141		0.0682**		-0.00404
		(0.0121)		(0.0321)		(0.0131)
employ_typ_3		0.00914		0.0126		0.00648
		(0.0136)		(0.0496)		(0.0236)
hukou	-0.0234**	-0.00854	-0.113***	-0.103***	0.000936	-0.000357
	(0.00923)	(0.0127)	(0.0141)	(0.0308)	(0.00986)	(0.0123)
hh_health	-0.0152*	-0.0214	-0.0220	-0.103***	-0.0404***	-0.0349**
	(0.00862)	(0.0146)	(0.0162)	(0.0341)	(0.0150)	(0.0173)
health_insurance	-0.0201	-0.0393*	0.0357	-0.0290	-0.00103	-0.00130**
	(0.0144)	(0.0239)	(0.0263)	(0.0441)	(0.000721)	(0.000663)
pension	-0.00293	0.0162	0.00969	0.0114	0.000624	-0.000582
	(0.0120)	(0.0176)	(0.0177)	(0.0296)	(0.00165)	(0.00211)
house_owner	-0.0458*	-0.0582	0.0672***	0.0647**	0.00454	-0.0303
	(0.0265)	(0.0494)	(0.0170)	(0.0275)	(0.0239)	(0.0291)
boy_number	-0.0756**	-0.0222	-0.0440	-0.000347	-0.0401	-0.0225
	(0.0319)	(0.0295)	(0.0441)	(0.0475)	(0.0253)	(0.0255)
girl_number	-0.0822**	-0.0288	-0.0582	0.0145	-0.0489*	-0.0337
	(0.0323)	(0.0335)	(0.0477)	(0.0540)	(0.0266)	(0.0271)
childage_04	0.0917**	0.0327	0.0720	0.122*	0.0332	0.0162
	(0.0358)	(0.0318)	(0.0510)	(0.0702)	(0.0281)	(0.0325)
childage_59	0.0621*	-0.00946	0.0645	0.0712	0.0491	0.0275
	(0.0345)	(0.0323)	(0.0498)	(0.0576)	(0.0305)	(0.0320)
childage_1014	0.0465	-0.0113	0.0827	0.105	0.0463*	0.0198
	(0.0338)	(0.0304)	(0.0515)	(0.0678)	(0.0278)	(0.0301)
childage_1519	0.0185	-0.0275	-0.00349	-0.00794	0.0238	0.0216
	(0.0325)	(0.0294)	(0.0513)	(0.0623)	(0.0273)	(0.0268)
age	0.00579*	-0.00313	0.00431	-0.0133*	0.000673	-0.00655
	(0.00316)	(0.00442)	(0.00303)	(0.00755)	(0.00236)	(0.00431)
age2	-2.50e-05	6.81e-05	-4.22e-05	0.000195**	-5.86e-06	8.62e-05*
	(3.36e-05)	(4.98e-05)	(2.83e-05)	(8.49e-05)	(2.47e-05)	(4.78e-05)
gender	0.0327***	0.0322***	0.0225	0.0118	0.0336***	0.0226*
	(0.00883)	(0.0122)	(0.0147)	(0.0284)	(0.00953)	(0.0119)
married	-0.0335**	-0.00862	-0.0729***	-0.0355	-0.000690	0.0266
	(0.0139)	(0.0216)	(0.0183)	(0.0342)	(0.0144)	(0.0240)
ccp_member	-0.0192*	-0.0233	-0.0410**	-0.0497	-0.0130	-0.0119
	(0.0106)	(0.0148)	(0.0174)	(0.0349)	(0.0110)	(0.0135)
yos	-0.00794***	-0.0102***	-0.00898***	-0.00130	-0.00589***	-0.00753***
	(0.00161)	(0.00213)	(0.00158)	(0.00352)	(0.00134)	(0.00195)
hh_credit	-0.104***	-0.0903***	-0.0181	0.0117	0.0976***	0.0789***
	(0.00872)	(0.0118)	(0.0130)	(0.0273)	(0.0145)	(0.0176)
Province FE	YES	YES	YES	YES	YES	YES
Constant	-0.0512	0.281***	-0.0442	0.419**	0.167**	0.348**
	(0.0742)	(0.104)	(0.0853)	(0.202)	(0.0799)	(0.172)
Observations	11,236	5,683	4,229	1,766	5,894	4,019

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(b) Saving rate and permanent income

VARIABLES	CHFS 2015 Urban (1) (sub-sample reg.)	CHFS 2015 Urban (2) (3 years avg. inc.)	CFPS 2014 Urban (3) (sub-sample reg.)	CFPS 2014 Urban (4) (3 years avg. inc.)	CHIP 2013 Urban (5) (sub-sample reg.)	CHIP 2013 Urban (6) (3 years avg. inc.)
hh_income_2	0.347***	0.0509**	0.255***	0.0313	0.0959***	0.0832***
	(0.0307)	(0.0207)	(0.0491)	(0.0577)	(0.0182)	(0.0117)
hh_income_3	0.512***	0.120***	0.308***	0.136**	0.169***	0.136***
	(0.0317)	(0.0255)	(0.0520)	(0.0547)	(0.0196)	(0.0130)
hh_income_4	0.617***	0.156***	0.472***	0.143*	0.175***	0.166***
	(0.0300)	(0.0223)	(0.0521)	(0.0784)	(0.0174)	(0.0127)
hh_income_5	0.819***	0.190***	0.616***	0.173***	0.257***	0.234***
	(0.0319)	(0.0265)	(0.0620)	(0.0606)	(0.0194)	(0.0153)
hh_size	-0.00777	0.0283***	0.0152	-0.00168	-0.00144	0.00240
	(0.00949)	(0.00933)	(0.0321)	(0.0194)	(0.0123)	(0.00499)
YDratio	0.0391	-0.0479	-0.0291	-0.217	-0.0594	-0.0826*
	(0.0681)	(0.100)	(0.172)	(0.181)	(0.0693)	(0.0476)
ODratio	0.121	0.0466		-0.0533	-0.0343	0.0392**
	(0.0923)	(0.0633)	(0.0947)	(0.0770)	(0.0176)	
employed	0.0648*	0.113***	0.0347	0.0126	0.0208	0.0419***
	(0.0363)	(0.0235)	(0.0751)	(0.0467)	(0.0353)	(0.0117)
hukou	-0.00258	0.00891	-0.0655*	-0.0630*	0.000555	-0.00260
	(0.0172)	(0.0220)	(0.0391)	(0.0361)	(0.0173)	(0.0111)
hh_health	-0.0379**	-0.0556***	-0.0484	-0.0415	-0.0459**	-0.0424***
	(0.0182)	(0.0210)	(0.0651)	(0.0381)	(0.0219)	(0.00878)
health_insurance	0.0102	0.000288	0.00430	0.0293	-0.00100	-0.000165
	(0.0255)	(0.0350)	(0.0464)	(0.0678)	(0.00131)	(0.000617)
pension	0.0234	0.0360	0.0917***	0.0633	0.000445	-0.000399
	(0.0186)	(0.0253)	(0.0353)	(0.0443)	(0.00249)	(0.000723)
house_owner	-0.0938	-0.150**	0.0433	0.0812	-0.0296	-0.00581
	(0.0839)	(0.0662)	(0.0331)	(0.0536)	(0.0376)	(0.0252)
boy_number	-0.00880	-0.100*	-0.0447	-0.0631	0.00134	-0.0451**
	(0.0425)	(0.0514)	(0.0742)	(0.0793)	(0.0328)	(0.0178)
girl_number	-0.0143	-0.126**	-0.0624	-0.0945	-0.0113	-0.0485**
	(0.0448)	(0.0541)	(0.0834)	(0.0864)	(0.0362)	(0.0200)
childage_04	0.00847	0.0697	0.0211	-0.000895	0.00919	0.0165
	(0.0463)	(0.0463)	(0.0962)	(0.0834)	(0.0353)	(0.0238)
childage_59	-0.0197	0.0720	0.0607	0.153*	0.00537	0.0494**
	(0.0453)	(0.0507)	(0.0847)	(0.0891)	(0.0356)	(0.0235)
childage_1014	-0.0269	0.0788	0.0540	0.109	0.00506	0.0525**
	(0.0434)	(0.0521)	(0.101)	(0.0948)	(0.0361)	(0.0238)
childage_1519	-0.0236	0.0249	0.0105	0.0698	-0.000359	0.0293
	(0.0426)	(0.0502)	(0.0838)	(0.0907)	(0.0307)	(0.0212)
age		-0.00787		0.0141		-0.00323
		(0.00606)		(0.00875)		(0.00222)
age2		9.53e-05		-8.97e-05		3.56e-05
		(6.53e-05)		(8.00e-05)		(2.23e-05)
gender	0.0228	-0.00320	0.0344	0.0498	0.0406**	0.0233**
	(0.0140)	(0.0180)	(0.0429)	(0.0425)	(0.0160)	(0.0103)
married	-0.0568*	0.0385	-0.129**	-0.0340	0.0125	-0.00566
	(0.0303)	(0.0238)	(0.0560)	(0.0524)	(0.0382)	(0.0135)
ccp_member	-0.0360	-0.0207	-0.0282	-0.0528	-0.0199	-0.0130
	(0.0223)	(0.0332)	(0.0428)	(0.0498)	(0.0156)	(0.00855)
yos	-0.00718***	0.00289	-0.00751	-0.00507	-0.00587**	-0.00576***
	(0.00242)	(0.00270)	(0.00498)	(0.00420)	(0.00257)	(0.00134)
hh_credit	-0.0975***	-0.0853***	-0.0148	0.0115	0.0970***	0.0644***
	(0.0178)	(0.0172)	(0.0337)	(0.0356)	(0.0189)	(0.0125)
Province FE	YES	YES	YES	YES	YES	YES
Constant	0.0554	0.279*	-0.133	-0.326	0.207**	0.259***
	(0.105)	(0.159)	(0.118)	(0.286)	(0.0820)	(0.0727)
Observations	3,901	6,849	965	1,030	2,149	5,879

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(c) Robust check: nonentrepreneurs and older households

	CHFS 2015 Urban		CFPS 2014 Urban		CHIP 2013 Urban	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	(excld. entrep.)	(age $\geq$ 60)	(excld. entrep.)	(age $\geq$ 60)	(excld. entrep.)	(age $\geq$ 60)
hh_income_2	0.436***	0.695***	0.226***	0.841***	0.0909***	0.0769***
	(0.0636)	(0.0436)	(0.0235)	(0.112)	(0.0125)	(0.0253)
hh_income_3	0.448***	0.990***	0.291***	1.164***	0.148***	0.122***
	(0.0810)	(0.0430)	(0.0245)	(0.114)	(0.0119)	(0.0304)
hh_income_4	0.864***	1.194***	0.426***	1.271***	0.184***	0.184***
	(0.0779)	(0.0505)	(0.0239)	(0.128)	(0.0133)	(0.0287)
hh_income_5	0.910***	1.445***	0.599***	1.412***	0.244***	0.274***
	(0.0902)	(0.0527)	(0.0240)	(0.140)	(0.0154)	(0.0351)
hh_size	-0.0649**	-0.0345**	-0.0336***	-0.0531	0.00122	-0.00758
	(0.0280)	(0.0147)	(0.0105)	(0.0377)	(0.00516)	(0.0160)
YDratio	-0.224	-0.611***	-0.180**	0.504	-0.0793	-0.336
	(0.372)	(0.231)	(0.0846)	(0.389)	(0.0499)	(0.227)
ODratio	-0.486	0.107***	0.0292	0.323**	0.0448***	0.0122
	(0.329)	(0.0286)	(0.0462)	(0.144)	(0.0170)	(0.0394)
employed	0.117	0.181***	0.00277	0.193**	0.0355***	0.0892***
	(0.0746)	(0.0378)	(0.0214)	(0.0840)	(0.0120)	(0.0292)
hukou	0.0476	-0.0303	-0.112***	-0.141	-0.00145	0.0110
	(0.0544)	(0.0307)	(0.0193)	(0.107)	(0.0119)	(0.0278)
hh_health	0.0558	-0.104***	-0.0109	-0.0174	-0.0373***	-0.0546***
	(0.0671)	(0.0212)	(0.0179)	(0.0485)	(0.00864)	(0.0207)
health_insurance	-0.0409	0.00727	0.0465	-0.148	-0.000311	0.00239
	(0.0876)	(0.0721)	(0.0293)	(0.0895)	(0.000678)	(0.00193)
pension	0.127*	0.00395	-0.00722	-0.0806	-0.000471	0.00630
	(0.0724)	(0.0443)	(0.0180)	(0.165)	(0.000769)	(0.00440)
house_owner	-0.786***	0.0836	0.0411*	0.0578	0.00176	0.104**
	(0.177)	(0.115)	(0.0215)	(0.0937)	(0.0244)	(0.0443)
boy_number	-0.00417	-0.0741	0.0142	-0.0546	-0.0457**	-0.0784
	(0.0957)	(0.0878)	(0.0402)	(0.155)	(0.0181)	(0.0657)
girl_number	-0.0214	-0.0198	0.0116	-0.0772	-0.0476**	-0.0488
	(0.113)	(0.0941)	(0.0435)	(0.152)	(0.0202)	(0.0733)
childage_04	0.0703	0.0946	0.0249	-0.122	0.0122	0.0717
	(0.160)	(0.0994)	(0.0483)	(0.152)	(0.0251)	(0.0916)
childage_59	0.0681	0.123	0.0135	-0.0137	0.0476*	0.124
	(0.145)	(0.111)	(0.0484)	(0.147)	(0.0245)	(0.0903)
childage_1014	0.130	0.107	0.0312	0.0665	0.0546**	0.143
	(0.147)	(0.106)	(0.0503)	(0.139)	(0.0256)	(0.0891)
childage_1519	-0.0256	-0.0515	-0.0362	-0.000626	0.0291	-0.0161
	(0.103)	(0.0952)	(0.0475)	(0.158)	(0.0221)	(0.0778)
age	0.0282	0.00475	-0.0190	-0.00341	0.0480	
	(0.0268)	(0.00380)	(0.0816)	(0.00215)	(0.0294)	
age2	-0.000295	-3.58e-05	9.37e-05	3.47e-05	-0.000313	
	(0.000277)	(3.60e-05)	(0.000540)	(2.15e-05)	(0.000202)	
gender	-0.00232	0.0712***	0.0294	0.162**	0.0235**	0.0165
	(0.0555)	(0.0240)	(0.0186)	(0.0738)	(0.0106)	(0.0258)
married	-0.103	-0.100***	-0.0685***	-0.108	-0.00565	-0.0155
	(0.0900)	(0.0348)	(0.0231)	(0.0816)	(0.0133)	(0.0285)
ccp_member	-0.0305	0.0276	-0.0602***	-0.0385	-0.0125	0.00554
	(0.106)	(0.0349)	(0.0225)	(0.0597)	(0.00873)	(0.0207)
yos	-0.0264**	-0.00904**	-0.00764***	-0.0156**	-0.00662***	-0.00705**
	(0.0104)	(0.00353)	(0.00202)	(0.00721)	(0.00140)	(0.00312)
hh_credit	0.0184	-0.0864***	-0.00521	-0.0550	0.0709***	0.0907**
Province FE	YES	YES	YES	YES	YES	YES
	(0.0483)	(0.0226)	(0.0166)	(0.0668)	(0.0121)	(0.0358)
constant	0.394	-0.747***	-0.189*	0.231	0.263***	-1.563
	(0.667)	(0.141)	(0.105)	(2.989)	(0.0723)	(1.047)
Observations	5,243	5,380	4,228	1,633	5,555	1,182

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(d) Robust check: per capital income and alternative definition of saving rate

	CHFS 2015 Urban		CFPS 2014 Urban		CHIP 2013 Urban	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	(redef. s.r.)	(re-grouped)	(redef. s.r.)	(re-grouped)	(redef. s.r.)	(re-grouped)
hh_income_2	0.326***	0.347***	0.257***	0.208***	0.128***	0.0842***
	(0.0213)	(0.0287)	(0.0289)	(0.0610)	(0.0186)	(0.0193)
hh_income_3	0.529***	0.519***	0.349***	0.309***	0.222***	0.151***
	(0.0231)	(0.0273)	(0.0332)	(0.0657)	(0.0192)	(0.0201)
hh_income_4	0.739***	0.635***	0.533***	0.451***	0.289***	0.190***
	(0.0233)	(0.0273)	(0.0310)	(0.0591)	(0.0206)	(0.0203)
hh_income_5	1.208***	0.829***	0.858***	0.625***	0.431***	0.251***
	(0.0328)	(0.0290)	(0.0387)	(0.0600)	(0.0286)	(0.0188)
hh_size	-0.0266**	0.0661***	-0.0476***	0.0715**	0.00703	0.0441***
	(0.0105)	(0.0113)	(0.0154)	(0.0344)	(0.00837)	(0.0148)
YDratio	-0.0258	0.0511	-0.315***	-0.121	-0.0940	-0.0811
	(0.0938)	(0.0795)	(0.107)	(0.183)	(0.0779)	(0.0781)
ODratio	0.0213	0.162**	0.0327	-0.0950	0.0853***	-0.0993
	(0.0766)	(0.0750)	(0.0613)	(0.0678)	(0.0308)	(0.0947)
employed	0.0799***	0.116***	0.0116	0.116***	0.0597***	0.0153
	(0.0194)	(0.0296)	(0.0302)	(0.0296)	(0.0182)	(0.0422)
hukou	-0.0442**	-0.000267	-0.185***	-0.0320	-0.0134	-0.00190
	(0.0188)	(0.0187)	(0.0288)	(0.0414)	(0.0223)	(0.0144)
hh_health	-0.0281	-0.0481**	-0.0171	-0.0623	-0.0674***	-0.0299
	(0.0175)	(0.0216)	(0.0271)	(0.0619)	(0.0140)	(0.0191)
health_insurance	0.000201	-0.00388	0.0699*	-0.0478	0.000656	-0.00103
	(0.0318)	(0.0290)	(0.0378)	(0.0540)	(0.00151)	(0.00140)
pension	-0.00141	0.000259	-0.0111	0.120***	-0.00162	0.000593
	(0.0210)	(0.0219)	(0.0233)	(0.0374)	(0.00134)	(0.00228)
house_owner	-0.122**	-0.0510	0.0546*	0.102**	-0.0118	-0.0215
	(0.0612)	(0.0849)	(0.0310)	(0.0414)	(0.0436)	(0.0393)
boy_number	-0.115**	-0.0191	-0.0165	0.0188	-0.0738**	-0.0145
	(0.0478)	(0.0407)	(0.0515)	(0.0846)	(0.0294)	(0.0373)
girl_number	-0.142***	-0.0156	-0.0233	0.0206	-0.0841**	-0.0230
	(0.0514)	(0.0428)	(0.0586)	(0.0866)	(0.0335)	(0.0387)
childage_04	0.119**	0.0293	0.0716	-0.0368	0.0130	0.00650
	(0.0520)	(0.0464)	(0.0641)	(0.106)	(0.0392)	(0.0451)
childage_59	0.0713	0.0125	0.0501	0.0155	0.0612	0.0104
	(0.0575)	(0.0454)	(0.0629)	(0.103)	(0.0389)	(0.0419)
childage_1014	0.0762	-0.0146	0.0799	0.00798	0.0703*	0.0160
	(0.0528)	(0.0425)	(0.0667)	(0.108)	(0.0392)	(0.0426)
childage_1519	0.0241	-0.0203	-0.000193	-0.00355	0.0366	0.00482
	(0.0520)	(0.0432)	(0.0613)	(0.0913)	(0.0340)	(0.0334)
age	0.00519	-0.0235	0.00925*	0.00111	-0.00435	0.0216
	(0.00782)	(0.0278)	(0.00533)	(0.000894)	(0.00379)	(0.0280)
age2	-1.35e-05	0.000396	-7.66e-05	0.0173	4.27e-05	-0.000300
	(8.23e-05)	(0.000367)	(5.21e-05)	(0.0409)	(3.75e-05)	(0.000370)
gender	0.0341*	0.0154	0.0471*	-0.0109	0.0354**	0.0410***
	(0.0189)	(0.0137)	(0.0249)	(0.0631)	(0.0159)	(0.0152)
married	-0.0422*	-0.000399	-0.109***	-0.0363	0.00144	-0.0187
	(0.0237)	(0.0389)	(0.0320)	(0.0499)	(0.0201)	(0.0335)
ccp_member	0.00961	-0.0272	-0.0828***	-0.0157	-0.0186	-0.0177
	(0.0293)	(0.0193)	(0.0280)	(0.0670)	(0.0145)	(0.0151)
yos	-0.0136***	-0.00508**	-0.0119***	-0.0133***	-0.0112***	-0.00564**
	(0.00275)	(0.00249)	(0.00291)	(0.00511)	(0.00234)	(0.00230)
hh_credit	-0.202***	-0.101***	-0.0475**	-0.0131	0.0731***	0.120***
	(0.0187)	(0.0180)	(0.0233)	(0.0348)	(0.0205)	(0.0230)
Province FE	YES	YES	YES	YES	YES	YES
Constant	0.0845	-0.0323	-0.148	1.683	0.399***	-0.293
	(0.179)	(0.532)	(0.138)	(1.247)	(0.123)	(0.523)
Observations	11,236	3,901	4,228	965	5,894	2,149

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(e) Check validity of the theoretical model

VARIABLES	CHFS 2015 Urban (1) hh_savingrate	(2) hh_(prev. w/curr. inc.)	(3) hh_savingrate	CFPS 2014 Urban (4) hh_savingrate	(5) hh_(prev. w/curr. inc.)	(6) hh_savingrate
hh_(prev. w/curr. inc.)			-0.00226***			-0.00288***
			(0.000436)			(0.000769)
hh_income_2	0.286***	-6.333***	0.290***	0.226***	-4.402***	0.224***
	(0.0173)	(1.200)	(0.0220)	(0.0235)	(1.123)	(0.0276)
hh_income_3	0.442***	-10.23***	0.434***	0.291***	-5.906***	0.277***
	(0.0176)	(1.029)	(0.0240)	(0.0245)	(1.171)	(0.0275)
hh_income_4	0.556***	-12.15***	0.534***	0.426***	-6.838***	0.418***
	(0.0181)	(1.126)	(0.0218)	(0.0239)	(1.099)	(0.0291)
hh_income_5	0.756***	-15.00***	0.719***	0.599***	-8.061***	0.580***
	(0.0199)	(1.355)	(0.0272)	(0.0240)	(1.174)	(0.0285)
hh_size	-0.0235***	-0.205	-0.0201**	-0.0336***	-0.488*	-0.0338***
	(0.00734)	(0.327)	(0.00833)	(0.0105)	(0.282)	(0.0112)
YDratio	0.0205	-2.921	-0.0429	-0.180**	1.766	-0.145
	(0.0748)	(2.679)	(0.0807)	(0.0846)	(2.168)	(0.101)
ODratio	0.0528	0.854	0.0409	0.0292	2.372	0.0455
	(0.0594)	(1.974)	(0.0625)	(0.0462)	(1.464)	(0.0507)
employed	0.0595***	1.290	0.0563***	0.00277	1.771**	0.00814
	(0.0147)	(1.111)	(0.0165)	(0.0214)	(0.736)	(0.0242)
hukou	-0.0181	1.192	-0.0198	-0.112***	-0.0575	-0.106***
	(0.0127)	(0.838)	(0.0194)	(0.0193)	(0.697)	(0.0232)
hh_health	-0.0168	-1.298*	-0.00522	-0.0109	-0.775	-0.0109
	(0.0144)	(0.705)	(0.0162)	(0.0179)	(0.764)	(0.0204)
health_insurance	0.00298	-1.973**	0.0112	0.0465	-1.112	0.0674*
	(0.0227)	(0.883)	(0.0284)	(0.0293)	(1.100)	(0.0366)
pension	-0.00220	1.332*	0.00602	-0.00722	-0.597	-0.0192
	(0.0144)	(0.728)	(0.0187)	(0.0180)	(0.585)	(0.0206)
house_owner	-0.0882**	6.180***	-0.179***	0.0411*	3.493***	0.0749***
	(0.0412)	(1.618)	(0.0487)	(0.0215)	(0.949)	(0.0273)
boy_number	-0.0718**	-0.00158	-0.0975**	0.0142	0.0599	-0.00257
	(0.0341)	(0.959)	(0.0452)	(0.0402)	(0.874)	(0.0476)
girl_number	-0.0923**	-0.385	-0.110**	0.0116	-0.507	-0.00326
	(0.0374)	(1.144)	(0.0466)	(0.0435)	(0.999)	(0.0512)
childage_04	0.0855**	1.940	0.122***	0.0249	1.043	0.0205
	(0.0379)	(1.255)	(0.0430)	(0.0483)	(1.008)	(0.0563)
childage_59	0.0487	1.567	0.101**	0.0135	1.882	0.0256
	(0.0427)	(1.258)	(0.0490)	(0.0484)	(1.234)	(0.0531)
childage_1014	0.0605	1.106	0.0904*	0.0312	0.527	0.0385
	(0.0389)	(1.321)	(0.0460)	(0.0503)	(1.142)	(0.0579)
childage_1519	0.0176	0.604	0.0538	-0.0362	0.770	-0.0363
	(0.0381)	(1.228)	(0.0446)	(0.0475)	(1.085)	(0.0559)
age	0.00428	-0.00525	0.00471	0.00475	0.403***	0.0121***
	(0.00483)	(0.218)	(0.00566)	(0.00380)	(0.115)	(0.00424)
age2	-1.06e-05	0.00188	-2.57e-05	-3.58e-05	-0.00260**	-0.000101***
	(5.21e-05)	(0.00244)	(5.97e-05)	(3.60e-05)	(0.00106)	(3.83e-05)
gender	0.0197	0.0995	0.0145	0.0294	-1.200*	0.0385*
	(0.0139)	(0.874)	(0.0174)	(0.0186)	(0.656)	(0.0220)
married	-0.0200	-1.114	-0.0323	-0.0685***	0.886	-0.0737***
	(0.0175)	(1.606)	(0.0245)	(0.0231)	(0.828)	(0.0270)
ccp_member	0.00140	-2.071**	-0.0103	-0.0602***	0.229	-0.0614***
	(0.0221)	(0.992)	(0.0326)	(0.0225)	(0.765)	(0.0227)
yos	-0.00793***	0.274***	-0.00717***	-0.00764***	0.283**	-0.00737***
	(0.00196)	(0.0974)	(0.00230)	(0.00202)	(0.115)	(0.00236)
hh_credit	-0.108***	2.797***	-0.0980***	-0.00521	1.653**	0.00170
	(0.0130)	(0.848)	(0.0146)	(0.0166)	(0.749)	(0.0182)
Province FE	YES	YES	YES	YES	YES	YES
Constant	-0.110	17.19***	0.0745	-0.189*	4.290	-0.386***
	(0.116)	(5.056)	(0.140)	(0.105)	(4.401)	(0.131)
Observations	6,849	6,849	6,849	3,235	3,235	3,235

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

Table B.5: The poor are more likely to face liquidity constraints

(a) CHFS 2015 Urban

VARIABLES	(1) (f.a. measure)	(2) (f.a. measure) (3 years avg. inc.)	(3) (cc measure)	(4) (cc measure) (3 years avg. inc.)
hh_income_1	0.100***	0.120***	0.270***	0.174***
	(0.0200)	(0.0239)	(0.00906)	(0.0143)
hh_income_2	0.0900***	0.0685***	0.229***	0.160***
	(0.0185)	(0.0202)	(0.00886)	(0.0137)
hh_income_3	0.0545***	0.0448**	0.188***	0.119***
	(0.0171)	(0.0215)	(0.00922)	(0.0148)
hh_income_4	0.0204	0.00889	0.130***	0.0723***
	(0.0164)	(0.0199)	(0.00969)	(0.0149)
hh_size	0.0116**	0.0102	3.34e-05	-0.0273***
	(0.00505)	(0.00652)	(0.00599)	(0.00682)
YDratio	-0.0653	-0.133*	-0.229***	-0.246***
	(0.0583)	(0.0746)	(0.0570)	(0.0677)
ODratio	-0.0210	-0.0202	0.120***	0.134**
	(0.0390)	(0.0501)	(0.0463)	(0.0550)
employed	-0.0178	-0.0172	0.0100	-0.0246
	(0.0121)	(0.0149)	(0.0150)	(0.0168)
hukou	-0.0830***	-0.0769***	-0.0883***	-0.0949***
	(0.0114)	(0.0144)	(0.0123)	(0.0142)
hh_health	0.0661***	0.0775***	0.00966	0.0341***
	(0.0109)	(0.0134)	(0.0111)	(0.0122)
health_insurance	-0.0331*	-0.0579**	-0.0275	-0.0117
	(0.0182)	(0.0240)	(0.0168)	(0.0232)
pension	-0.122***	-0.134***	-0.0702***	-0.0844***
	(0.0131)	(0.0168)	(0.0134)	(0.0154)
house_owner	-0.0219	0.0493	-0.0428	-0.0591
	(0.0489)	(0.0606)	(0.0385)	(0.0649)
age	0.00323	0.0122**	-0.0167***	0.000654
	(0.00355)	(0.00509)	(0.00386)	(0.00556)
age2	-1.60e-05	-0.000102*	0.000264***	8.48e-05
	(3.84e-05)	(5.47e-05)	(4.39e-05)	(6.19e-05)
gender	-0.0257**	-0.0136	0.0128	0.00724
	(0.0106)	(0.0137)	(0.00933)	(0.0127)
married	-0.0461***	-0.0512**	0.00962	0.00916
	(0.0159)	(0.0213)	(0.0159)	(0.0214)
ccp_member	-0.0104	0.0461	-0.0346**	-0.00545
	(0.0139)	(0.0285)	(0.0162)	(0.0263)
yos	-0.0191***	-0.0165***	-0.0273***	-0.0271***
	(0.00158)	(0.00196)	(0.00174)	(0.00214)
Province FE	YES	YES	YES	YES
Observations	11,236	6,849	11,236	6,849

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(b) CFPS 2014 Urban and CHIP 2013 Urban

VARIABLES	(1) (f.a. measure)	(2) (f.a. measure) (3 years avg. inc.)	(3) (f.a. measure)	(4) (f.a. measure) (3 years avg. inc.)
hh_income_1	0.232***	0.195***	0.135***	0.145***
	(0.0300)	(0.0349)	(0.0303)	(0.0272)
hh_income_2	0.204***	0.136**	0.104***	0.105***
	(0.0270)	(0.0532)	(0.0248)	(0.0246)
hh_income_3	0.0927***	0.128***	0.0412*	0.0441**
	(0.0274)	(0.0327)	(0.0232)	(0.0204)
hh_income_4	0.0553**	0.0547*	0.0658***	0.0569**
	(0.0268)	(0.0303)	(0.0219)	(0.0234)
hh_size	0.00322	0.00866	0.0178***	0.0196***
	(0.00701)	(0.00814)	(0.00639)	(0.00631)
YDratio	-2.40e-05	-0.0263	-0.00654	-0.00428
	(0.0489)	(0.0642)	(0.0409)	(0.0403)
ODratio	-0.0455	-0.0547	-0.0574**	-0.0594**
	(0.0370)	(0.0426)	(0.0277)	(0.0273)
employed	-0.0212	-0.0124	-0.0208	-0.0200
	(0.0215)	(0.0248)	(0.0149)	(0.0149)
hukou	-0.0822***	-0.0935***	0.0251	0.0231
	(0.0236)	(0.0247)	(0.0182)	(0.0184)
hh_health	0.0720***	0.0741***	0.0374**	0.0358**
	(0.0196)	(0.0230)	(0.0162)	(0.0163)
health_insurance	-0.0314	-0.0451	0.000329	0.000341
	(0.0282)	(0.0373)	(0.000554)	(0.000547)
pension	-0.00806	-0.0127	-0.00342	-0.00328
	(0.0193)	(0.0242)	(0.00235)	(0.00232)
house_owner	-0.0280	-0.0180	-0.0388	-0.0314
	(0.0203)	(0.0244)	(0.0368)	(0.0357)
age	0.00651*	0.00988**	-0.00440	-0.00436
	(0.00341)	(0.00492)	(0.00285)	(0.00280)
age2	-5.63e-05*	-7.73e-05*	3.43e-05	3.53e-05
	(3.42e-05)	(4.52e-05)	(2.71e-05)	(2.66e-05)
gender	0.00215	0.0384*	-0.00287	-0.00460
	(0.0181)	(0.0220)	(0.0132)	(0.0132)
married	-0.0412*	-0.0683**	0.0211	0.0187
	(0.0251)	(0.0308)	(0.0178)	(0.0176)
ccp_member	-0.0247	-0.0431*	0.0128	0.0132
	(0.0253)	(0.0253)	(0.0139)	(0.0139)
yos	-0.00483**	-0.00748***	-0.00162	-0.00107
	(0.00231)	(0.00269)	(0.00195)	(0.00197)
Province FE	YES	YES	YES	YES
Observations	4,387	3,016	4,907	4,898

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

Table B.6: The effect of liquidity constraints on household saving rate

	CHFS 2015 Urban				CFPS 2014 Urban	
VARIABLES	(1) (cc measure)	(2) (cc measure)	(3) (f.a. measure)	(4) (f.a. measure)	(5) (f.a. measure)	(6) (f.a. measure)
credit × 2015	-0.278***	-0.114***	-0.263***	-0.212***	-0.264**	-0.240**
	(0.0775)	(0.0357)	(0.0673)	(0.0440)	(0.118)	(0.230)
credit	-0.0942*	-0.194***	0.0192	-0.0486*	0.0114	-0.0688***
	(0.0548)	(0.0421)	(0.0335)	(0.0252)	(0.0398)	(0.0251)
2015	-0.279***	-0.130***	0.0783	-0.0190	-0.244***	-0.159***
	(0.0368)	(0.0257)	(0.0494)	(0.0357)	(0.0437)	(0.0252)
income	2.13e-06***		1.93e-06***		9.27e-06***	
	(3.44e-07)		(2.84e-07)		(1.04e-06)	
loginc		0.756***		0.778***		0.798***
		(0.0159)		(0.0181)		(0.00912)
hh_size	0.160***	-0.0408***	0.161***	-0.0421***	0.0794***	-0.131***
	(0.0175)	(0.0121)	(0.0188)	(0.0153)	(0.0201)	(0.0102)
YDratio	-0.0321	-0.147	-0.0937	-0.197	-0.242	-0.475***
	(0.220)	(0.166)	(0.240)	(0.188)	(0.169)	(0.105)
ODratio	-0.304***	-0.00651	-0.736***	-0.224*	-0.0933	0.158***
	(0.0679)	(0.0527)	(0.159)	(0.116)	(0.0775)	(0.0459)
employed	0.185***	0.186***	0.236***	0.204***	0.103***	0.0794***
	(0.0411)	(0.0307)	(0.0441)	(0.0341)	(0.0350)	(0.0209)
hukou	0.0755**	-0.243***	0.0714*	-0.132***	-0.0817**	-0.321***
	(0.0367)	(0.0297)	(0.0405)	(0.0326)	(0.0369)	(0.0232)
hh_health	-0.180***	-0.0564**	-0.181***	-0.0539*	-0.0880***	-0.0112
	(0.0348)	(0.0277)	(0.0377)	(0.0292)	(0.0310)	(0.0184)
health_insurance	0.0278	-0.0900**	0.0601	-0.0523	-0.0136	-0.0429
	(0.0602)	(0.0422)	(0.0601)	(0.0422)	(0.0526)	(0.0300)
pension	0.248***	-0.0102	0.266***	-0.00377	-0.0313	-0.0161
	(0.0381)	(0.0287)	(0.0399)	(0.0311)	(0.0395)	(0.0242)
house_owner	0.0249	0.0960***	-0.0313	0.0146	0.156***	0.0554*
	(0.0487)	(0.0363)	(0.0527)	(0.0369)	(0.0501)	(0.0295)
boy_number	-0.164**	-0.155**	-0.199**	-0.151**	-0.0608*	-0.00537
	(0.0747)	(0.0604)	(0.0778)	(0.0639)	(0.0358)	(0.0228)
girl_number	-0.152*	-0.150**	-0.172**	-0.151**	-0.0185	-0.0118
	(0.0829)	(0.0668)	(0.0871)	(0.0712)	(0.0400)	(0.0241)
childage_04	-0.0540	0.148**	-0.0440	0.144*	-0.133**	0.135***
	(0.0895)	(0.0708)	(0.0970)	(0.0779)	(0.0545)	(0.0340)
childage_59	-0.0474	0.141**	-0.0122	0.150*	-0.116**	0.120***
	(0.0908)	(0.0702)	(0.0986)	(0.0767)	(0.0522)	(0.0328)
childage_1014	-0.0915	0.110	-0.0632	0.132	-0.0390	0.145***
	(0.0948)	(0.0749)	(0.102)	(0.0819)	(0.0589)	(0.0383)
childage_1519	-0.0888	0.0194	-0.0768	0.0236	-0.101**	-0.00102
	(0.0889)	(0.0699)	(0.0938)	(0.0742)	(0.0393)	(0.0262)
age	-0.00552	0.0179***	0.0198	0.0160	0.0183**	0.0325***
	(0.00830)	(0.00683)	(0.0134)	(0.0104)	(0.00858)	(0.00490)
age2	0.000127	-4.37e-05	-0.000165	-2.41e-05	-0.000144*	-0.000224***
	(7.78e-05)	(6.45e-05)	(0.000143)	(0.000111)	(7.92e-05)	(4.49e-05)
gender	0.00717	-0.0195	0.0641	0.0198	0.0496**	0.0690***
	(0.0380)	(0.0290)	(0.0406)	(0.0325)	(0.0208)	(0.0150)
married	-0.150***	-0.278***	-0.121**	-0.248***	-0.0555	-0.0759***
	(0.0513)	(0.0384)	(0.0574)	(0.0448)	(0.0435)	(0.0248)
ccp_member	-0.122	-0.0565	-0.141	-0.105	-0.0446	-0.127***
	(0.0911)	(0.0604)	(0.0970)	(0.0642)	(0.0508)	(0.0314)
yos	0.0170***	-0.0146***	0.0106*	-0.0253***	-0.00382	-0.0161***
	(0.00466)	(0.00362)	(0.00541)	(0.00436)	(0.00370)	(0.00227)
Province FE	YES	YES	YES	YES	YES	YES
Constant	-0.787***	-8.220***	-1.294***	-8.309***	-4.58e-06**	-8.785***
	(0.260)	(0.254)	(0.329)	(0.300)	(1.91e-06)	(0.211)
Observations	9,588	9,588	7,024	7,024	7,209	7,209
R-squared	0.231	0.558	0.243	0.576	0.237	0.579

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

Table B.7: The effect of income inequality on aggregate household saving rate

	CHFS 2015 Urban		CFPS 2014 Urban		CHIP 2013 Urban	
VARIABLES	county_savingrate	county_savingrate	county_savingrate	county_savingrate	county_savingrate	county_savingrate
county gini	0.198*	0.176**	0.135	0.139	0.0188	0.00636
	(0.115)	(0.0710)	(0.413)	(0.412)	(0.126)	(0.125)
logcounty_pc_inc	0.322***	0.329***	0.00946*	0.00967*	0.0962***	0.107***
	(0.0178)	(0.0201)	(0.00535)	(0.00599)	(0.0223)	(0.0229)
county_YDratio		0.353		-0.0237		0.277
		(0.330)		(0.0967)		(0.177)
county_ODratio		0.583**		0.0354		-0.104
		(0.283)		(0.0744)		(0.151)
Constant	-3.093***	-3.290***	0.0301	0.0284	-0.637***	-0.761***
	(0.178)	(0.225)	(0.173)	(0.189)	(0.241)	(0.251)
Province FE	YES	YES	YES	YES	YES	YES
Observations	353	353	334	334	212	212
R-squared	0.521	0.527	0.227	0.227	0.152	0.169

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

Table B.8: The estimates of marginal propensity to consume out of transitory income

(a) CHFS Urban

VARIABLES	Quintile 1 hh_consump	Quintile 2 hh_consump	Quintile 3 hh_consump	Quintile 4 hh_consump	Quintile 5 hh_consump
hh_per_income	0.788***	0.610***	0.374*	0.541***	0.223**
	(0.112)	(0.135)	(0.191)	(0.125)	(0.0889)
hh_tran_income	0.788***	0.611***	0.364*	0.507***	0.178*
	(0.112)	(0.135)	(0.191)	(0.127)	(0.0952)
hh_netasset	0.00258	0.00267	0.00552***	0.00378*	0.00755***
	(0.00178)	(0.00189)	(0.00184)	(0.00203)	(0.00108)
hh_size	1,845***	2,256***	1,928**	-455.1	7,738**
	(520.1)	(746.1)	(964.9)	(1,603)	(2,975)
YDratio	3,356	3,752	7,838	728.0	-53,291***
	(5,460)	(5,902)	(9,017)	(9,438)	(19,647)
ODratio	255.8	1,937	-6,332**	-8,972	-6,840
	(1,664)	(3,839)	(2,968)	(6,439)	(13,406)
employed	-835.8	-3,095	-5,642	-3,028	3,586
	(1,155)	(2,406)	(3,619)	(4,439)	(6,465)
hukou	1,115	1,618	3,557	-5,127	4,924
	(1,020)	(1,945)	(2,923)	(4,383)	(6,983)
hh_health	468.2	600.7	5,046*	5,281	13,002
	(1,252)	(1,763)	(2,840)	(4,557)	(8,132)
health_insurance	818.5	-5,301*	5,876	10,938**	-37,742
	(2,548)	(2,983)	(5,334)	(4,575)	(26,362)
pension	107.5	-1,709	-354.5	8,811***	15,302
	(1,528)	(2,093)	(5,229)	(3,223)	(13,844)
house_owner	963.7	6,240	10,855**		-11,075
	(2,840)	(10,370)	(4,182)		(25,768)
age	-309.7	-405.6	-2,064**	-1,498	550.6
	(311.6)	(758.6)	(949.9)	(922.6)	(2,172)
age2	2.048	1.537	15.82*	12.43	-9.408
	(2.641)	(6.461)	(8.723)	(7.890)	(19.29)
gender	-2,579*	-2,514	-1,020	1,485	18,535
	(1,311)	(2,731)	(3,287)	(4,034)	(12,512)
married	3,402**	3,223	6,112**	6,899	4,335
	(1,501)	(2,730)	(2,465)	(7,658)	(8,798)
ccp_member	2,003	7,668	-9,152	-10,602*	1,014
	(3,108)	(4,754)	(10,935)	(6,198)	(12,878)
yos	201.1	393.1	-329.7	1,042**	19.43
	(233.7)	(388.4)	(400.1)	(443.0)	(1,212)
Province FE	YES	YES	YES	YES	YES
Constant	3,139	16,225	51,066*	48,290	6,421
	(10,747)	(23,458)	(27,895)	(39,625)	(83,059)
Observations	294	436	479	476	488
R-squared	0.650	0.220	0.222	0.228	0.238

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(b) CFPS Urban

VARIABLES	Quintile 1 hh_consump	Quintile 2 hh_consump	Quintile 3 hh_consump	Quintile 4 hh_consump	Quintile 5 hh_consump
hh_per_income	0.874***	0.347**	0.831***	0.486**	0.393***
	(0.102)	(0.169)	(0.198)	(0.208)	(0.126)
hh_tran_income	0.927***	0.361*	0.739***	0.477**	0.336***
	(0.0985)	(0.187)	(0.203)	(0.206)	(0.125)
hh_netasset	0.00410**	0.00453	0.00425	0.00102	0.00764**
	(0.00159)	(0.00360)	(0.00343)	(0.00196)	(0.00369)
hh_size	879.4*	1,423**	1,621	2,557**	-840.1
	(489.1)	(672.4)	(1,050)	(989.1)	(2,085)
YDratio	-1,503	-4,587	1,425	8,354	23,179
	(3,648)	(5,285)	(6,333)	(9,386)	(17,154)
ODratio	776.4	-4,393	1,852	-6,132	-6,818
	(1,785)	(3,251)	(5,194)	(6,606)	(19,803)
employed	-692.5	-1,254	3,645	5,278	8,038
	(978.3)	(2,234)	(3,030)	(3,654)	(6,518)
hukou	831.1	4,431**	3,471	7,554**	14,384**
	(1,014)	(1,708)	(2,131)	(3,533)	(6,118)
hh_health	-716.9	-363.0	2,192	-429.5	-2,001
	(749.6)	(1,979)	(2,406)	(3,483)	(5,962)
health_insurance	-94.29	624.9	38.74	-9,354	-8,738
	(1,407)	(2,759)	(2,618)	(6,569)	(8,984)
pension	1,007	-1,382	-1,900	251.4	9,983*
	(1,016)	(2,184)	(2,435)	(2,971)	(5,519)
house_owner	-2,441**	-3,162	-4,790	860.8	-3,897
	(1,155)	(2,573)	(3,417)	(3,256)	(7,581)
age	-310.8	-565.2	-1.083	-979.6	-918.4
	(325.9)	(416.9)	(551.5)	(831.7)	(1,486)
age2	1.740	4.446	-0.978	11.08	11.59
	(2.735)	(3.716)	(4.861)	(7.028)	(12.55)
gender	905.9	-2,282	-849.3	-5,015	-1,582
	(784.4)	(1,980)	(2,426)	(3,648)	(6,571)
married	-1,966*	5,344**	-40.79	9,584**	14,149**
	(1,070)	(2,071)	(2,749)	(4,458)	(5,956)
ccp_member	1,470	3,282	2,242	2,882	-461.7
	(1,513)	(2,064)	(3,024)	(3,230)	(6,547)
yos	146.4	121.6	768.0***	347.7	133.1
	(129.8)	(190.0)	(265.2)	(337.8)	(770.2)
Province FE	YES	YES	YES	YES	YES
Constant	22,369**	19,787	11,894	16,818	29,539
	(10,159)	(14,757)	(21,953)	(28,023)	(47,455)
Observations	359	489	546	572	494
R-squared	0.557	0.216	0.226	0.162	0.204

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

(c) CHIP Urban

VARIABLES	Quintile 1 hh_consump	Quintile 2 hh_consump	Quintile 3 hh_consump	Quintile 4 hh_consump	Quintile 5 hh_consump
hh_per_income	0.613***	0.623***	0.510***	0.346***	0.233***
	(0.0287)	(0.0707)	(0.0827)	(0.0871)	(0.0806)
hh_tran_income	0.563***	0.502***	0.453***	0.308***	0.0167
	(0.0732)	(0.111)	(0.107)	(0.101)	(0.111)
hh_netasset	0.00258	-0.00342	0.00283	0.000505	0.00129
	(0.00338)	(0.00523)	(0.00306)	(0.00378)	(0.00356)
hh_size	400.5	556.1	679.0	1,246*	1,052
	(264.7)	(414.3)	(559.5)	(683.8)	(1,234)
YDratio	940.3	1,880	8,397**	-2,443	11,196
	(1,524)	(2,736)	(3,752)	(5,187)	(8,620)
ODratio	-475.2	-3,291**	-1,227	-5,937*	-16,334**
	(991.9)	(1,594)	(2,141)	(3,374)	(6,964)
employed	-1,458**	-1,731*	-2,714*	-2,834	-4,635
	(638.1)	(1,042)	(1,384)	(1,827)	(4,941)
hukou	-1,325**	80.83	1,145	-366.2	6,422
	(626.6)	(998.8)	(1,439)	(1,870)	(4,076)
hh_health	777.7*	2,542**	1,204	5,521***	1,689
	(458.1)	(997.9)	(1,389)	(1,882)	(4,341)
health_insurance	148.1	6.603	61.08	102.8	72.48
	(115.5)	(56.30)	(92.83)	(63.91)	(178.4)
pension	-32.62	-8.11	14.03	43.11	879.7
	(78.36)	(156.2)	(41.50)	(169.0)	(549.1)
house_owner	-649.5	119.0	-143.1	2,579	323.7
	(992.3)	(2,177)	(2,818)	(4,149)	(8,969)
age	-70.71	-173.2	146.3	134.9	1,110**
	(133.4)	(224.0)	(257.9)	(411.2)	(449.7)
age2	-0.0399	1.116	-1.415	-1.741	-7.981*
	(1.322)	(2.171)	(2.526)	(3.944)	(4.668)
gender	-348.6	-165.2	-1,901	-2,882*	-5,462**
	(636.8)	(861.6)	(1,275)	(1,575)	(2,609)
married	502.3	724.7	102.3	-2,231	-4,509
	(632.4)	(1,475)	(1,539)	(2,792)	(4,368)
ccp_member	-265.5	297.8	1,392	160.3	1,592
	(634.0)	(848.8)	(1,003)	(1,608)	(2,260)
yos	108.2	339.4***	480.4***	850.1***	1,617***
	(86.61)	(123.5)	(172.5)	(224.7)	(489.5)
Province FE	YES	YES	YES	YES	YES
Constant	8,524**	4,150	-634.0	12,734	1,985
	(3,964)	(8,733)	(8,572)	(15,165)	(22,473)
Observations	1,214	1,205	1,236	1,236	1,240
R-squared	0.363	0.122	0.098	0.074	0.237

Notes: 1. Robust standard errors (cluster at county level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1; 2. Standard errors in parentheses are bootstrapped 500 times for bootstrapped median regression.

APPENDIX C

DOES THE COLLEGE ADMISSION POLICY IMPROVE HIGH SCHOOL EDUCATION IN CHINA'S POOR COUNTIES?

C.1 Figures

Figure C.1: First-tier universities admission rate

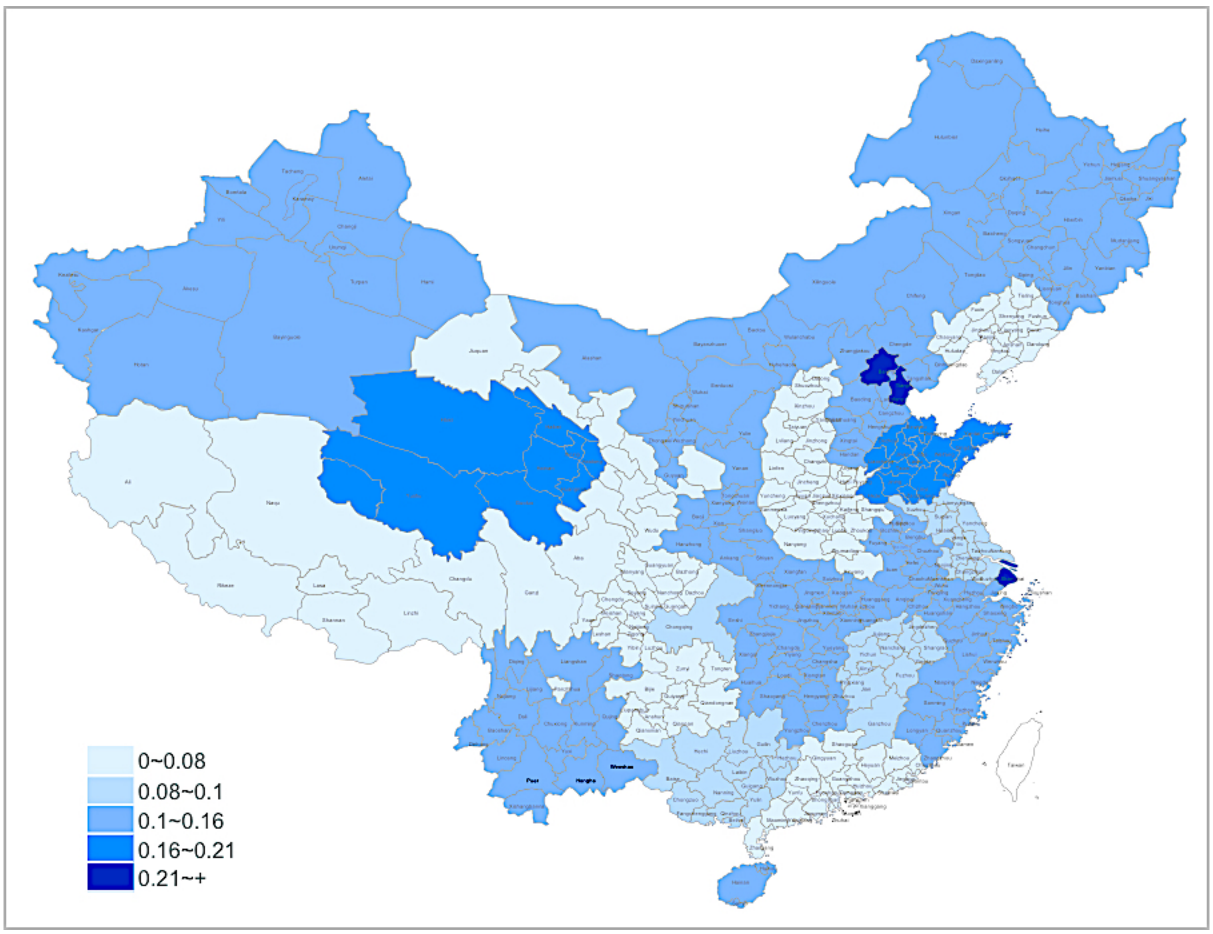
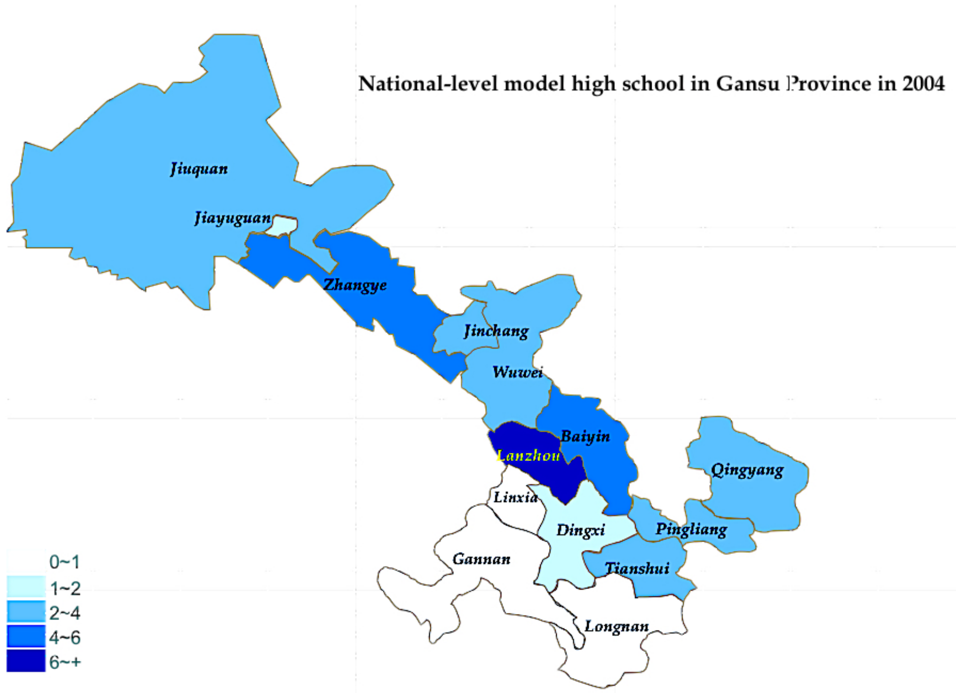


Figure C.2: National-level model high school in Gansu: 2004 and 2011



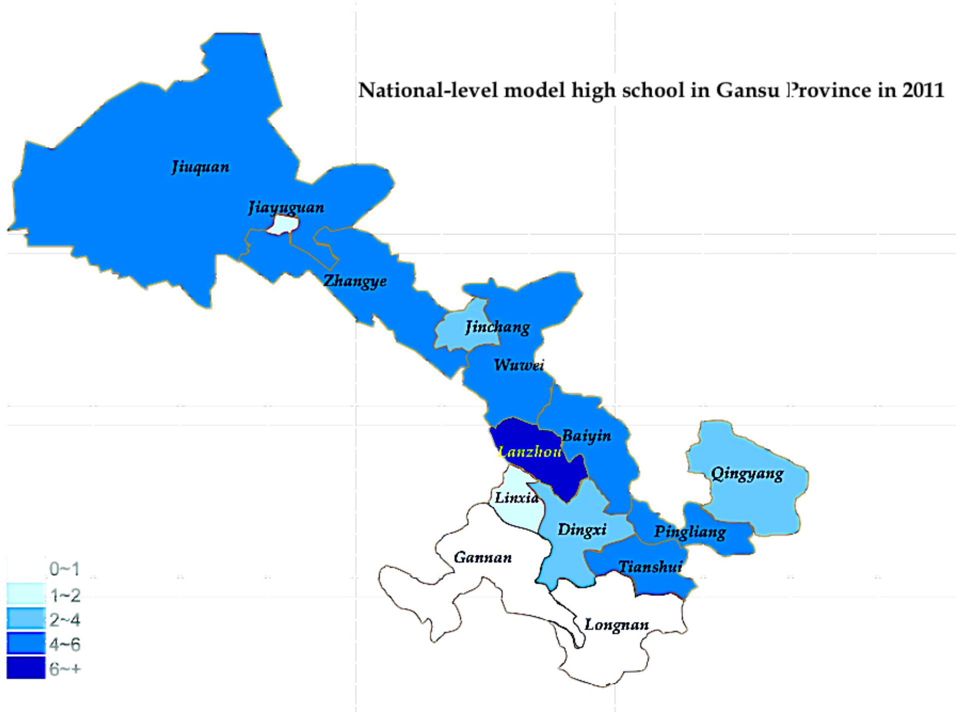


Figure C.3: Government expenditure per student on education in Gansu province

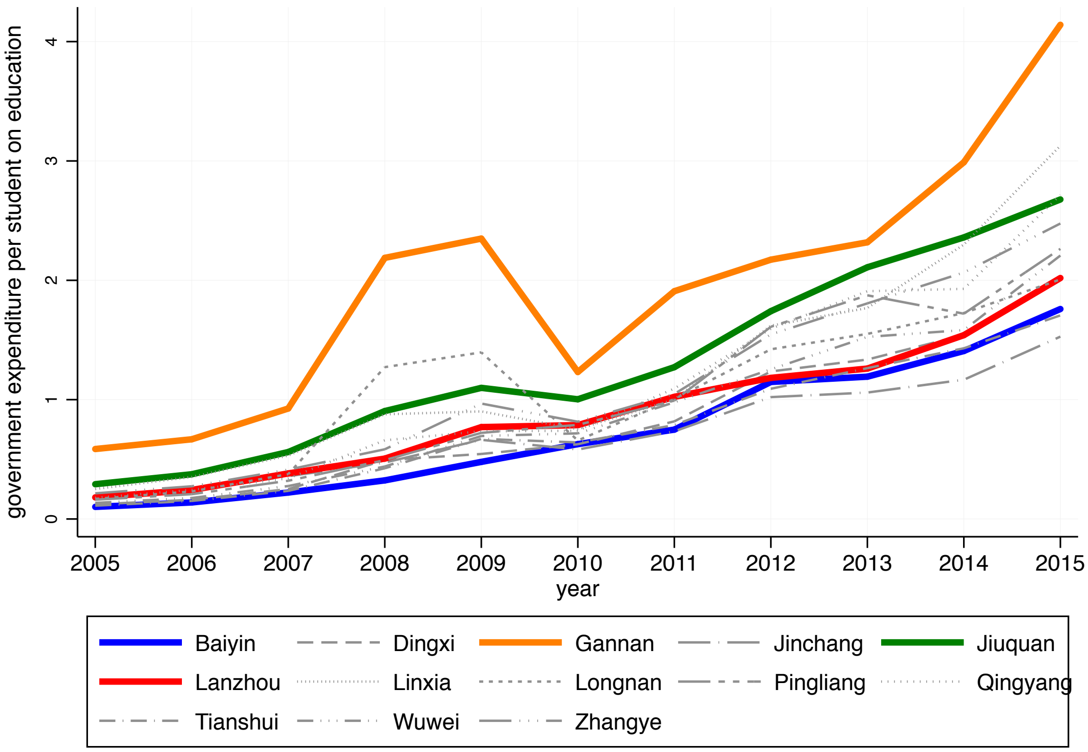


Figure C.4: Average entry rate of treatment and comparison counties group

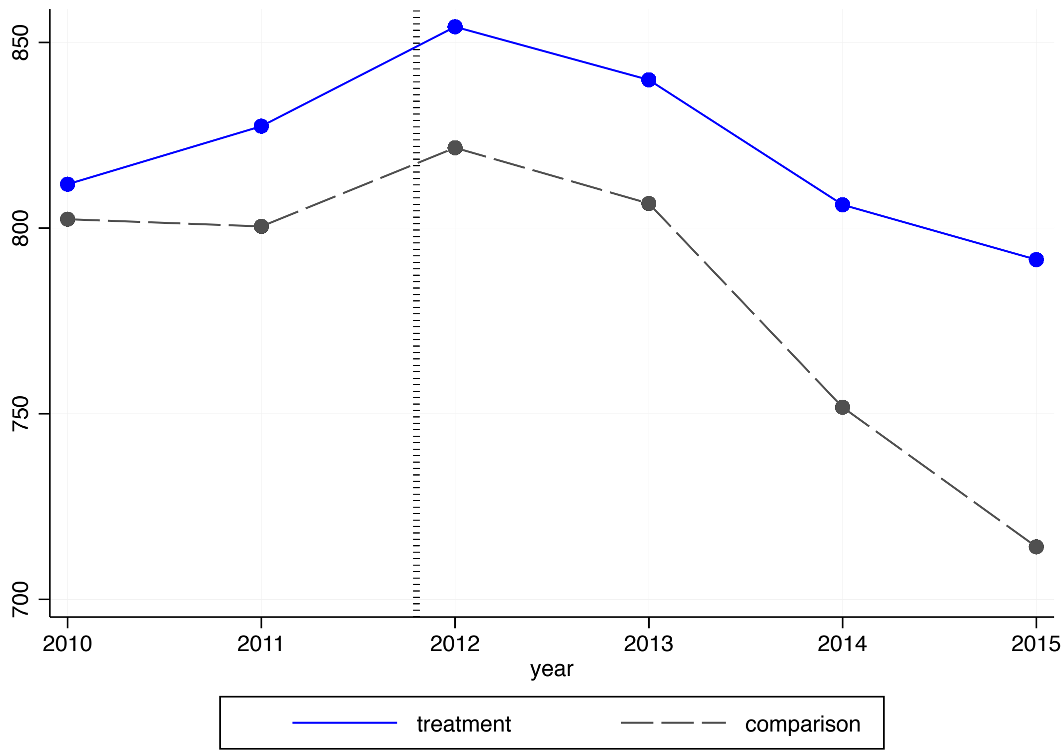


Figure C.5: Average enrollment rate of treatment and comparison counties group

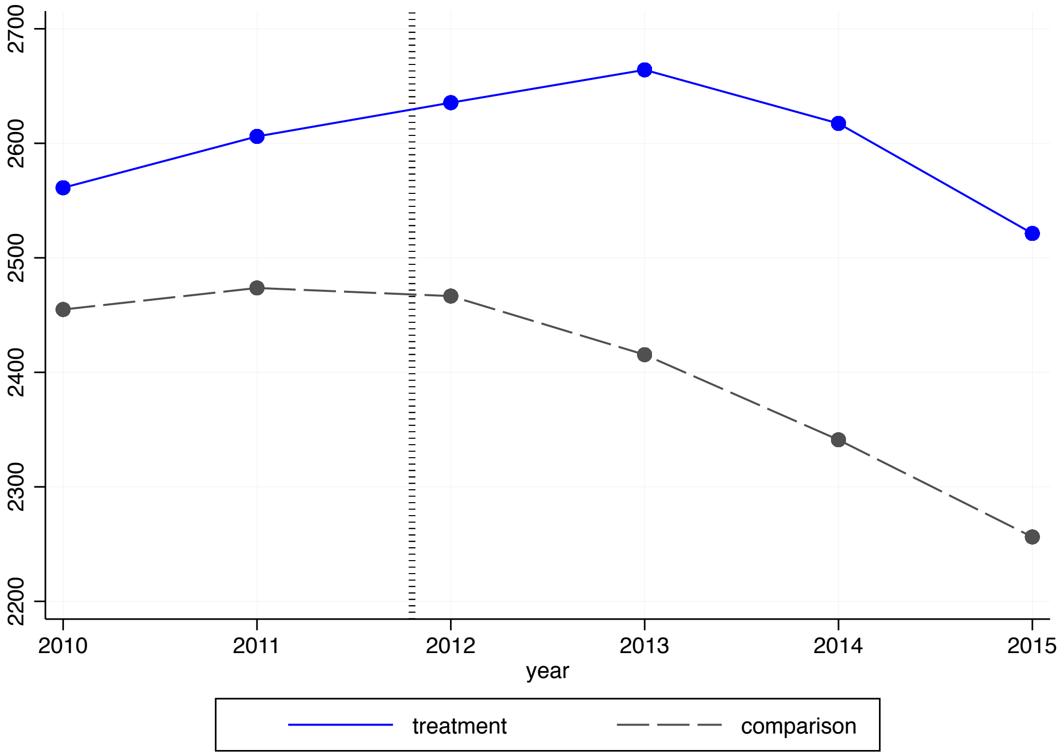


Figure C.6: Propensity score histogram by treatment and T-test by matching status

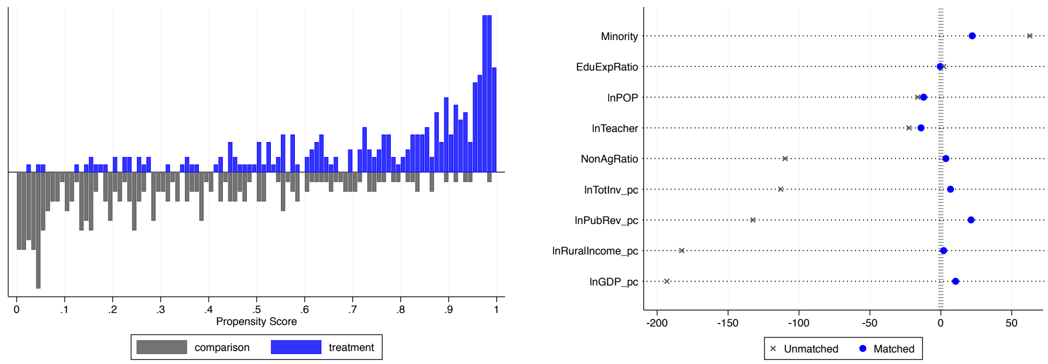
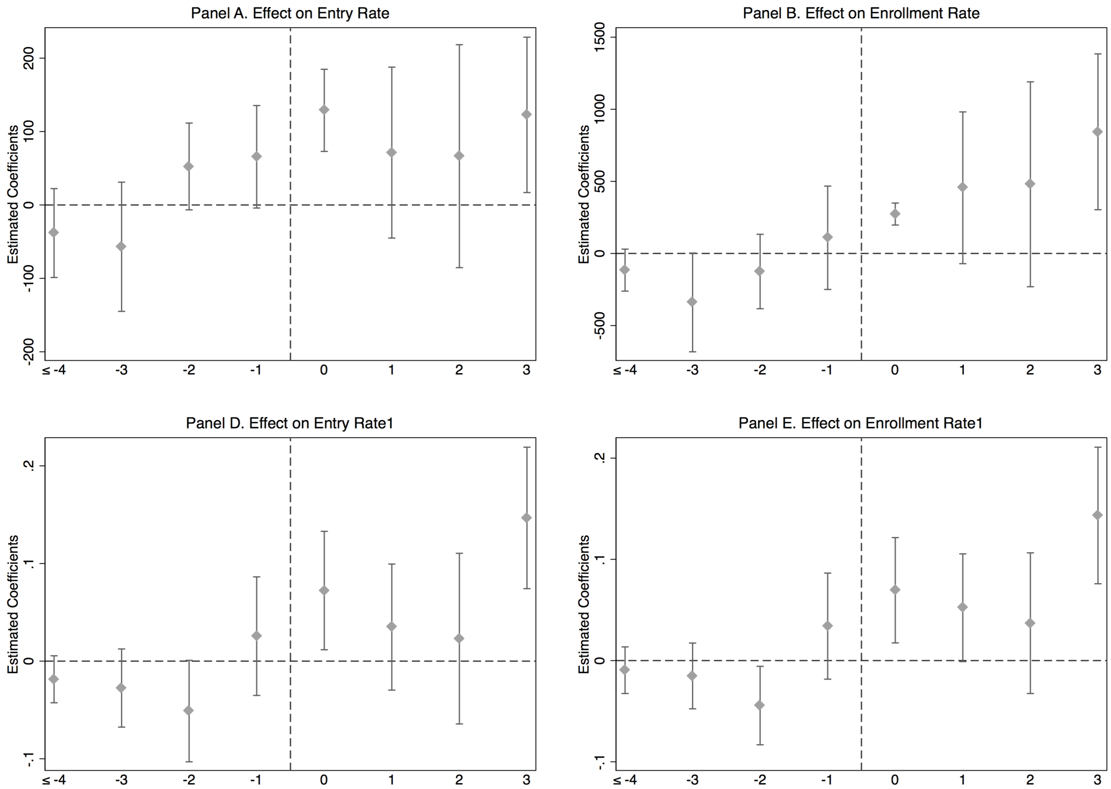


Figure C.7: Estimated dynamic effect of admission policy



C.2 Tables

Table C.1: Summary statistics: outcomes (mean value)

	2005-2011		2012-2015		(After2012-Before2012)		(Policy - Non-Policy)
	Policy	*Non-Policy*	*Policy*	*Non-Policy*	*Policy*	*Non-Policy*	*Policy - Non-Policy*
Entry rate	758.5	839.57	822.98	773.53	64.48	-66.04	130.52
Enrollment rate	2223.61	2509.8	2551.55	2407.92	327.94	-101.88	429.82
Graduation rate	694.92	802.21	927.37	872.61	232.45	70.4	162.05
Entry rate1	0.343	0.515	0.452	0.568	0.109	0.053	0.056
Enrollment rate1	0.324	0.499	0.451	0.566	0.127	0.067	0.060
Graduation rate1	0.922	0.952	0.983	0.988	0.061	0.036	0.025
Observations	406	196	232	112	638	318	956

Table C.2: Baseline regression results

(a) Entry rate 1

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
Policy ×After	99.07***	84.29	122.0***	224.9***
	(23.63)	(55.76)	(35.60)	(46.03)
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	812	812
R^2	0.755	0.806	0.735	0.815

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(b) Enrollment rate 1

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
Policy ×After	316.6***	244.2	223.7	586.2***
	(59.31)	(156.7)	(142.5)	(169.4)
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	812	812
R^2	0.785	0.841	0.782	0.862

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(c) Entry rate 2

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
Policy ×After	0.0654***	0.0539*	0.0757***	0.129***
	(0.0184)	(0.0295)	(0.0108)	(0.0323)
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	812	812
R^2	0.854	0.886	0.939	0.957

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(d) Enrollment rate 2

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
Policy ×After	0.0731***	0.0569*	0.0587***	0.119***
	(0.0188)	(0.0273)	(0.0125)	(0.0384)
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	812	812
R^2	0.862	0.898	0.950	0.968

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

Table C.3: Placebo test

	(1) Entry 1	(2) Enroll 1	(3) Entry 2	(4) Enroll 2
Policy ×Placebo	75.86	172.3	0.0450	0.0288
	(46.01)	(163.6)	(0.0297)	(0.0319)
Observations	602	602	602	602
R^2	0.436	0.492	0.461	0.496

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

Table C.4: Estimation on dynamic Effect and test for parallel trends assumption

	(1) Entry 1	(2) Enroll 1	(3) Entry 2	(4) Enroll 2
Policy ×Y2008	-38.26	-115.5	-0.0186	-0.0095
	(36.81)	(88.57)	(0.0146)	(0.0140)
Policy ×Y2009	-56.94	-339.6	-0.0275	-0.0152
	(53.50)	(208.1)	(0.0243)	(0.0197)
Policy ×Y2010	62.58	-125.0	-0.0511	-0.0341
	(35.93)	(156.9)	(0.0316)	(0.0236)
Policy ×Y2011	75.62	108.9	0.0356	0.0247
	(42.43)	(217.9)	(0.0369)	(0.0442)
Policy ×Y2012	128.9***	273.5***	0.0729***	0.0594***
	(34.02)	(46.38)	(0.0184)	(0.0190)
Policy ×Y2013	71.34	455.6	0.0943	0.0820
	(70.78)	(320.0)	(0.0640)	(0.0683)
Policy ×Y2014	66.49	479.9	0.0912	0.0833
	(92.32)	(432.0)	(0.0797)	(0.0828)
Policy ×Y2015	122.7*	843.9**	0.142**	0.149**
	(64.33)	(328.5)	(0.0598)	(0.0643)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	946	946
R^2	0.808	0.836	0.887	0.899

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

Table C.5: Heterogeneous policy regression results

(a) Heterogeneous policy effects on entry rate

	1%-25% (1)	25%-50% (2)	50%-75% (3)	75%-100% (4)
Policy ×After	89.28	64.18	162.0***	123.7***
	(52.32)	(55.12)	(39.80)	(15.03)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	462	484	473	473
R^2	0.905	0.890	0.809	0.913

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(b) Heterogeneous policy effects on entry rate 1

	1%-25% (1)	25%-50% (2)	50%-75% (3)	75%-100% (4)
Policy ×After	0.0635	0.0346	0.0808***	0.0725***
	(0.517)	(0.364)	(0.0135)	(0.0106)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	462	484	473	473
R^2	0.939	0.939	0.888	0.948

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(c) Heterogeneous policy effects on enrollment rate

	1%-25%	25%-50%	50%-75%	75%-100%
	(1)	(2)	(3)	(4)
Policy ×After	240.7	226.6	500.7**	485.4***
	(185.6)	(152.2)	(170.6)	(62.13)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	462	484	473	473
R^2	0.923	0.932	0.811	0.940

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(d) Heterogeneous policy effects on enrollment rate 1

	1%-25% (1)	25%-50% (2)	50%-75% (3)	75%-100% (4)
Policy ×After	0.0674	0.0410	0.0840***	0.0940***
	(0.0617)	(0.0331)	(0.0134)	(0.00947)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	462	484	473	473
R^2	0.943	0.951	0.889	0.956

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(e) Heterogeneous policy effects on graduation rate

	1%-25% (1)	25%-50% (2)	50%-75% (3)	75%-100% (4)
Policy ×After	96.89	75.59	153.6	194.8***
	(94.95)	(61.19)	(105.3)	(30.47)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	462	484	473	473
R^2	0.893	0.895	0.802	0.904

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

Table C.6: Underlying mechanism regression results

(a) Student-teacher ratio

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
Policy $\times$After	-1.593	-1.327	1.681	1.624
	(1.014)	(1.021)	(1.420)	(1.304)
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	812	812
R^2	0.844	0.847	0.899	0.901

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).

(b) Education expenditure (public finance)

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
Policy ×After	0.108*	0.0761	-0.110	-0.113
	(0.0575)	(0.0616)	(0.0674)	(0.0745)
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Prefecture-Year FE	Y	Y	Y	Y
Observations	946	946	812	812
R^2	0.978	0.979	0.989	0.989

Notes: 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 2. Controls include GDP *per capita* (in logarithm), education expenditure (in logarithm), and number of teachers (in logarithm).