

THREE ESSAYS ON
THE LABOR MARKET IN CHINA

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by ZIXIN LIU
Dr. Peter Mueser, Dissertation Supervisor

MAY 2021

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THREE ESSAYS ON
THE LABOR MARKET IN CHINA

presented by Zixin Liu,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Peter Mueser

Professor Cory Koedel

Professor Saku Aura

Professor Oded Gurantz

DEDICATION

This dissertation is dedicated to my grandpa and grandma in the heaven. They accompanied with me in my childhood, encouraged me to be an optimistic, honest, and hard-working person, and wished me to obtain a doctorate degree when I grew up. This dissertation was inspired by their stories. To my beloved parents, who always provide unconditional supports to me. To my advisor, Dr. Peter Mueser, who gave me a lot of valuable advice in both research and daily life. To my friends, who I shared my happiness and sadness with, and who helped me and encouraged me when I was in difficulties.

ACKNOWLEDGEMENTS

I am extremely grateful for my advisor, Dr. Peter Mueser, for his valuable advice, continuous support and patience in my PhD research. I would also give special thanks to my thesis committee members, Dr. Cory Koedel, Dr. Saku Aura, and Dr. Oded Gurantz, for providing many insightful suggestions to my dissertation. Moreover, I would like to express gratitude to all the professors who taught or worked with me in my PhD period, for the knowledge and opportunities they brought to me that help me develop deeper understanding in economics. Finally, I would like to thank my parents for their mental and material supports to me. Without their understanding and encouragement, it would be impossible for me to complete my study.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLESv

LIST OF FIGURESx

ABSTRACT xiii

CHAPTER

1. Impact of the Cultural Revolution on Intergenerational Mobility in China 1

 Introduction 1

 Literature Review 4

 Background 9

 Data 13

 Methods 16

 Results 22

 Extensions 35

 Conclusion..... 38

 Appendix 49

 Reference..... 80

2. Impact of the College Expansion on Skilled Labor Market in China.....86

 Introduction 86

 Background 89

Data	98
Empirical Models	101
Main Results.....	104
Conclusion.....	119
Appendix	139
Reference.....	156
3. The Impact of College Expansion on Internal Migration in China	160
Introduction	160
Literature Review	161
Data and Background	168
Models.....	179
Results.....	183
Robustness Checks	191
Conclusion.....	195
Appendix	213
Reference.....	221
VITA.....	230

LIST OF TABLES

Chapter 1

Table 1.1 Descriptive Statistics of the Weighted CHIP Sample.....	41
Table 1.2 Measures of Affected Years in the Cultural Revolution.....	42
Table 1.3 Estimated Coefficients of the Impact of Affected Years in the CR on Education	43
Table 1.4 Estimated Coefficients of the Impact of the CR on Education by Area of Birth	44
Table 1.5 Estimated Coefficients for the CR Impacts on Multi-Generational Educational Transmission.....	45
Table 1.6 Decomposition of the Effect of Father's Schooling on College Degree Attainment.....	46

Chapter 2

Table 2.1 Annual Increase Rate for Number of Employed Workers and Number of College Graduate Workers.....	121
Table 2.2 Annual Increase Rate for Number of Employed Workers and Number of College Graduate Workers (Urban).....	122
Table 2.3 Descriptive Statistics of Sample in CGSS 2003-2015 (Urban)	123
Table 2.4 Statistics for Labor Market Variables in CGSS 2003-2015 (Urban).....	124
Table 2.5 Regression Results of the College Expansion on Annual Income.....	125
Table 2.6 Regression Results of the College Expansion on Unemployment	126
Table 2.7 Regression Results of the College Expansion on Task Scores of First Job....	127

Chapter 3

Table 3.1 Summary Statistics on the CLDS 2012 Sample	197
Table 3.2 Summary Statistics for Migrants by Migration Reason and by Migration Cohort	198
Table 3.3 Estimated Results of Equation (3.6) on Migration Decision for Ages 17-20.	199
Table 3.4 Estimated Results of Equation (3.6) on Migration Decision Within 3 Years After Graduation	200
Table 3.5 Marginal Effects on Migration Decision on Migration Decision For Ages 17-20 Using McFadden’s Conditional Logit Model.....	201
Table 3.6 Marginal Effects on Migration Decision Within 3 Years After Graduation Using McFadden’s Conditional Logit Model.....	202
Table 3.7 Calculated Chi-square for Models with Full Sample and Subsample in Table 3.5 and Table 3.6.....	203

Appendix Tables

Chapter 1

APTable 1.1 Estimated Intergenerational Educational Coefficients by Birth Cohort, Replicating Chen et al. (2015): Years of Schooling	54
APTable 1.2 Estimated Intergenerational Educational Coefficients by Birth Cohort, Replicating Chen et al. (2015): Rank of Schooling	55
APTable 1.3 Estimated Intergenerational Educational Correlations by Birth Cohort, Replicating Chen et al. (2015): Years of Schooling	56
APTable 1.4 Population Distribution by Region in the CHIP Sample and the Chinese Yearbook 2013.....	57

APTable 1.5 Household Distribution and Average Household Size of the 14 CHIP Provinces in CHIP Sample and the Chinese Yearbook 2013	58
APTable 1.6 Demographic Comparison between CHIP Sample and Chinese Population by Province	59
APTable 1.7 Demographic Comparison between CHIP Sample and Chinese Population by Province	60
APTable 1.8 Estimated Coefficients of the Impact of Affected Years in the CR on Education Controlling Province Fixed Effects	61
APTable 1.9 Robustness Check: Estimated Coefficients of Equation (1.1), Dropping Elder Sample.....	62
APTable 1.10 Estimated Marginal Effects of Equation (1.1), Using Logit and Probit Models.....	63
APTable 1.11 Estimated Results of Equation (1.1), Using Schooling Missing Pattern and Years of Father’s Schooling.....	64
APTable 1.12 Estimated Results of Equation (1.1), Using Sibling Sample	65
APTable 1.13 Estimated Results of Equation (1.1), Adding Family Fixed Effects (Sibling Sample)	66
APTable 1.14 Estimated Results of Equation (1.1), Changing School Entry Age.....	67
APTable 1.15 Estimated Coefficients Using All Sample, Urban Sample, and Sample of Urban Residents Who Have Children in the CHIP.....	68
APTable 1.16 Decomposition of the Effect of Father's Schooling on College Degree Attainment Using Gelbach (2016)	69

APTable 1.17 Occupation as the Measure of Father's Status: Estimated Coefficients of the Effect of Affected Years in the CR on High School and College Degree Attainments	70
APTable 1.18 Occupation as Measure of Father's Status: Estimated Coefficients of the Impact of the CR on Multi-Generational Mobility for High School and College Degree	71
APTable 1.19 Estimated Coefficients of the Effect of Mother's Schooling and Father's Schooling on High School and College Attainments: Based on Equations (1.7), (1.8) and (1.9).....	72
APTable 1.20 Sample Descriptive Statistics by the Answer of "Whether Took CEE" in the CHIP.....	73
Chapter 2	
APTable 2.1 Regression Results for College Premium within Industry and within Occupation	147
APTable 2.2 Regression Results Based on Equations (2.2) and (2.3).....	148
APTable 2.3 Regression Results Based on Equation (2.1) Using the Ratio between Provincial College Admission and Population	149
APTable 2.4 Placebo Test on the Impact of the College Expansion on Other Macro-Level Variables	150
APTable 2.5 Marginal Effect Estimates Using Logit Model on Unemployment.....	151
APTable 2.6 Estimates of Returns to Task Scores within Education.....	152
APTable 2.7 Estimated Results of the First Stage in the 2SLS Regression	153
APTable 2.8 Estimated Results of the Second Stage in the 2SLS Regression on Income	154

APTable 2.9 Estimated Results of the Second Stage in the 2SLS Regression on Unemployment.....	155
--	-----

Chapter 3

APTable 3.1 Variable Description.....	213
APTable 3.2 Regression Results for Non-Eastern Migration for Ages 17-20.....	215
APTable 3.3 Regression Results for Non-Eastern Migration in 3 Years After Graduation	216
APTable 3.4 Regression Results for Migration for Ages 17-20 Using the Number of Teachers as the Measure for College Expansion	217
APTable 3.5 Regression Results for Migration in 3 Years After Graduation Using the Number of Teachers as the Measure for College Expansion.....	218
APTable 3.6 Regression Results for Migration for Education Between Age 17 to 20 ..	219
APTable 3.7 Regression Results for Migration for Working in 3 Years After Graduation	220

LIST OF FIGURES

Chapter 1

Figure 1.1 Estimated Intergenerational Educational Coefficients in 1925-1985..... 47

Figure 1.2 Average Years of Schooling across Birth Year by Father's Education 48

Chapter 2

Figure 2.1 Number of Students Admitted to Regular and Adult 4-Year and 2- or 3-Year Higher Education Institutions in 1992-2015..... 128

Figure 2.2 Average Annual Increase Rate of College Admissions in 1998-2012..... 129

Figure 2.3 Gross College Enrollment Ratio and Aged 18-22 Population in 1992-2015 130

Figure 2.4 National Public Budget and Ratio between Number of Students and Number of Teachers in Regular Colleges in 1992-2015..... 131

Figure 2.5 National Public Budget and Ratio between Number of Students and Number of Teachers in Adult Colleges in 1992-2015 132

Figure 2.6 Number of Colleges by Study in 1992-2015..... 133

Figure 2.7 Number of Colleges by Control Agency in 1992-2015..... 134

Figure 2.8 Educational Distribution for Employed Workers in 1996-2015 135

Figure 2.9 Change of Proportions in College Workers in China from 1998 to 2012..... 136

Figure 2.10 Predicted College Premium and Number of Students Admitted to Colleges in 1998-2015 137

Figure 2.11 Proportions of College Graduates Among Employed and Unemployed Workers..... 138

Chapter 3

Figure 3.1 The Floating Population and Its Regional Distribution in China 204

Figure 3.2 Educational Distribution of Migrants in China	204
Figure 3.3 Inter-provincial and Intra-provincial Migration Rates by Age Group and by Education in 3 Time Periods.....	205
Figure 3.4 Distribution of Migration Reasons by Education and Age Group	206
Figure 3.5 Migration Rates and Hukou Change Rates by Age Group and by Education	207
Figure 3.6 Proportion of Different Situations of Migration and Hukou Change by Age Group and by Education	208
Figure 3.7 Distribution of Regional Type of Hukou Change by Destinations, Age Group, and Education.....	209
Figure 3.8 Distribution of Hukou Change Reason by Age Group and by Education.....	210
Figure 3.9 Total Number and Regional Distribution of College Admissions by Year...	211
Figure 3.10 Number of College Admissions by Region and Type in China	212

Appendix Figures

Chapter 1

APFigure 1.1 Estimated Coefficients of the Intergenerational Coefficients of APTable 1.2 (Rank of Schooling).....	74
APFigure 1.2 Estimated Coefficients of the Intergenerational Correlations of APTable 1.3 (Years of Schooling).....	75
APFigure 1.3 Estimated Coefficients of the Intergenerational Schooling Coefficients and Correlations for Urban and Rural Residents.....	76
APFigure 1.4 Estimates of the Intergenerational Schooling Coefficients and Correlations for Male and Female	77

APFigure 1.5 Estimates of the Intergenerational Schooling Coefficients and Correlations
for Urban-born and Rural-born Samples 78

APFigure 1.6 Number of People Who Reported Taking the CEE as Proportions of 19-
year-old Population and the Proportion of 19-year-old Population Registered for the CEE
by Year..... 79

THREE ESSAYS ON
THE LABOR MARKET IN CHINA

Zixin Liu

Dr. Peter Mueser, Dissertation Supervisor

ABSTRACT

This dissertation consists of three chapters studying the impact of different policies in China on its labor market. In particular, we consider the impact of the Cultural Revolution on intergenerational mobility, the impact of college expansion on earnings and unemployment, and the impact of college expansion on migration.

In the first chapter, we study the impacts of the Cultural Revolution on intergenerational and multi-generational educational mobility in China. We use a difference-in-difference method to show that the Cultural Revolution (CR) significantly reduced the advantage of having a more-educated father on a child's educational attainment. The impact of the CR on intergenerational mobility is identified by an index that measures for each individual the number of school years during which the CR restricted education access. The decline of the effect of father's educational level on children's college degree attainment is mediated through the likelihood of obtaining a high school degree, participating in the college entrance examination, and obtaining higher exam scores for those who take the exam. However, the Cultural Revolution did not fully eliminate the advantage of having a more-educated father on a child's educational

achievement, nor did it reduce the effect of grandfather's schooling on a grandchild's educational achievement.

In the second chapter, we study the short-term response of the labor market to an unprecedented expansion in the Chinese higher education system from 1999 to 2012 on labor market outcomes for young and older college graduates and non-college graduates. Using the number of provincial college admissions as a measure of college expansion, we identify the impacts of the college expansion on the college premium, unemployment, and skills used in first jobs. In the short run, the college expansion decreased the college premium and increased the likelihood of unemployment for new college graduates. Also, the college expansion reduced the cognitive skills used in college graduates' first jobs. The negative impact of the college expansion on labor outcomes is smaller for older college graduates. Our results are consistent with findings published in the 1970s focusing on the effects of the U.S college expansion.

In the third chapter, we study the impact of aggregate college admissions on inter-provincial migration in China for different age groups. Examining migration propensity, we find that the college expansion has a direct "enrollment effect" and a "competition effect" on the likelihood of inter-provincial migration. College-bound students are more likely to migrate at ages 17-20 as college admissions in outside provinces increase; and college graduates are more likely to migrate after graduation as the number of local new college graduates increases. In addition, we identify a negative impact of local college admissions on migration at ages 17-20, reflecting the improvement in local educational and labor market opportunities. We also use a conditional Logit model to consider the choice

of migration destination and identify how inter-regional differences in college growth affect the patterns of migration.

These three chapters provide multiple policy implications as well as evidence for labor economic theories and hypotheses as they relate to China's labor market.

Chapter 1

Impact of the Cultural Revolution on Intergenerational Mobility in China

1.1 Introduction

China's Cultural Revolution (CR) of 1966-1976 instituted major changes in the country's educational and economic system, including closing schooling at all levels for three years, suspending the college entrance exam for ten years, altering curricula, and expanding primary and secondary schools in rural areas. People were publicly labeled based on their family backgrounds¹ in order to achieve the stated egalitarian goals of the CR. Government policies were linked to group membership. Therefore, opportunities to obtain education were also altered dramatically across family backgrounds, in addition to the general changes in educational system in the CR.

The impact of the Cultural Revolution on educational opportunities, return to schooling, beliefs, etc. have been frequently studied. Meng and Gregory (2007) found that years of schooling decreased significantly for people in urban areas who received their

¹ As noted in Deng and Treiman (1997):

I. Good-class origins, also referred to as the “five red kinds”: 1. Revolutionary cadres; 2. Revolutionary army soldiers; 3. Revolutionary martyrs (the orphans of men who died in the revolutionary wars); 4. Pre-liberation industrial workers and their families; 5. Former poor and lower-middle peasant families; II. Middle-class origins: 1. Families of pre-Liberation peddlers and store clerks, etc.; 2. Former middle-peasant families; 3. Families of pre-liberation clerks, teachers, professionals, etc.; III. Bad-class origins: 1. Families of former capitalists; 2. Families of “rightists” (a label denoting those who were outspoken in the Hundred Flowers campaign of 1957); 3. Pre-liberation rich peasant families; 4. Families of “bad elements” (a label denoting “criminal” offenders); 5. Pre-liberation landlord families; 6. Families of counterrevolutionaries.

education during the CR, which is “about 2 to 3 times the effect of World War II on the education attainment”. However, the CR promoted educational attainment for children growing up in rural areas (Han, 2001). Deng and Treiman (1997) showed that the effects of the CR differed based on parental social status. In an analysis limited to households with both parents and children, they found that people from educated or cadre families lost their advantages in obtaining higher educational achievements during the CR. In an analysis based on surveys of several cities, Giles et al. (2019) confirmed that the impact of the CR on a child's education was negatively correlated with parental occupation. The CR also changed people's beliefs. Roland and Yang (2017) found that people who were affected by the CR were less likely to believe in the value of effort for their own achievement, but they invested more in their children. Although the CR decreased people's educational attainment, it had a small impact on the returns to schooling (Zhang et al., 2007).

Most of the previous literature that studies the impact of the CR on the intergenerational mobility in China has been limited to urban data (Zhou, 2014; Chen et.al, 2015; Giles et al., 2019), and none of the research that studies the educational mobility has analyzed the underlying mechanism. Also, there are few studies on the impact of the CR on multi-generational mobility. One exception is Xie and Zhang (2019), which, like this chapter, uses a representative sample and considers both inter- and multigenerational model. Although their finding on the effect of the CR on the relationship between parental class and child's education are consistent with ours, the structure of their analysis fails to identify the direct impacts the specific policies in the CR, focusing on overall effects on the CR generation.

In this chapter, we use a nationally representative sample to study the impact of the CR on intergenerational educational mobility in China. The impact of the CR on intergenerational mobility is identified by an index indicating the number of school years during which the CR restricted an individual's education access, based on the person's birth year.

We explore the level of intergenerational mobility during the CR, estimate the impact of the CR on children's high school degree attainment and college degree attainment by father's education. We also estimate the differential impact of the CR on intergenerational mobility by urban and rural areas and analyze the potential mechanisms by which the CR affected intergenerational mobility. Moreover, we estimate the impact of the CR for a father's generation on their children, and how this relates to the effects of grandfathers' education.

We extend the literature by providing new perspectives in studying the impact of the CR on intergenerational educational mobility. We use a sample which covers both urban and rural areas and test whether the impact of the CR on intergenerational mobility differed by areas. We use a variation of a difference-in-difference (DiD) strategy to identify impacts of educational disruptions associated with the CR on intergenerational and multi-generational educational mobility. The impact is identified through comparing educational attainments for people with different years of disruption by the CR and with different family backgrounds. We also explore the underlying mechanisms of the impact of the CR on the intergenerational educational mobility.

The chapter is organized as follows: Section 1.2 reviews the literature on intergenerational mobility; Section 1.3 introduces the background for this research; Section 1.4 describes the data used in this study; Section 1.5 describes the methods and models; Section 1.6 presents estimated results; Section 1.7 discusses the further extensions; and Section 8 concludes the chapter.

1.2 Literature Review

Intergenerational mobility measures the change of social status between subsequent generations within families. A lower intergenerational mobility rate indicates less opportunity for people from poor families to improve their social status during their lifetime and less possibility for people from rich families to move down. Thus, it is an important measure of social inequality across generations. In empirical studies, a reduced-form model which relates children's status with parents' status has been widely used:

$$y_{it+1} = \beta y_{it} + \varepsilon_{it}$$

the dependent variable y_{it+1} can be any measure of the social status (such as income, education level, or occupation) for a person born in $t + 1$ cohort from family i . The independent variable y_{it} is the socioeconomic status for the person's parent. The degree of association between child's status and parents' status, β , can thus be interpreted as intergenerational coefficient. Another way to measure intergenerational mobility is intergenerational correlation, which is the product between intergenerational coefficient and the ratio of the standard deviations between child and parental generation:

$$\rho = \beta \left(\frac{\sigma_1}{\sigma_0} \right)$$

the intergenerational correlation can thus remove the difference in cross-sectional inequality across generations. A large intergenerational coefficient or correlation indicates a low level of intergenerational mobility.

Many empirical studies have estimated the levels and trends of intergenerational mobility in various countries. Even within countries with similar socioeconomic features (such as the OECD countries), the mobility levels can be very different. Estimates of the recent intergenerational income elasticity for the U.S is 0.4-0.5 between father and son. This suggests that 1% increase in father's income increases son's income by about 0.4-0.5%. In contrast, the elasticities are below 0.3 for the Nordic countries, indicating a higher intergenerational mobility (Jantti et al. 2006). For developing countries, Hertz et al. (2008) show that the intergenerational correlations for the South American countries are all above 0.55, higher than any other countries in his study. In contrast, the intergenerational correlation for rural China is only 0.2.

The trend of intergenerational mobility across cohorts also differs by country. In the U.K, there was a decrease of mobility during 1950-1970 (Blanden and Machin, 2004). In Nordic countries, the mobility did not decrease in 1950-1965 (Bratberg et al., 2005). While Lee and Solon (2009) and Chetty et al. (2014) found little statistical evidence supporting any significant change of the intergenerational earning elasticity in the U.S for people born in 1952-1993. Chen et al. (2015) showed an inverted U-shape pattern of the intergenerational educational mobility in 20th century China. The level of mobility they estimated was low in the pre-Mao era (for people born before the mid-1940s), reached its highest level in the Mao-era (for those born in the mid-1940s and the mid-1960s), and returned to a low level in the post-Mao era (for those born after the mid-1960s). Differences

of the estimated intergenerational income elasticity can be explained by the variations in return to human capital, policies and demographic structure across countries and time Solon (2004). Corak (2013) suggests intergenerational mobility is also negatively correlated with cross-sectional inequality.

Understanding the underlying mechanism behind the intergenerational coefficient is another important topic of this literature. The economic theory started from Becker (1976). In their model, people make decisions regarding their own consumption and the investment in the human capital of their children. Their model suggests that intergenerational earning elasticity depends on heritability of natural endowment and return to human capital. Becker and Tomes (1986) argued that when there is no constraint on investment, the transmission degree would solely depend on heritability of endowment. Thus, child's status is associated with parents' status mainly through nature (such as heritability of traits) and nurture (such as investment on child's human capital). External variables, such as return to education, policy, also play roles in the transmission process through affecting the decision or level for a household to invest on next generation.

The mechanisms of how parental status plays a role in affecting child's status can be categorized into nature (such as transmission of genes, abilities) and nurture (such as investment on child's human capital). Some studies suggest that nature is more important than nurture in intergenerational transmission process. Mazumder (2008) and Björklund et al. (2005) contribute the similarity between sibling's earning into nature and nurture, and both found that genetic factors are more important than shared environmental factors in explaining the sibling earning correlation. Clark (2015) focused on the changes of the percentages of elite families with rare surnames in the upper tail of social status distribution

to infer the underlying intergenerational transmission pattern. He showed that the intergenerational coefficients of the underlying competence between two generations in almost all countries were all between 0.7-0.9 and kept stable over time. He further argued that the observed low intergenerational mobility in all countries across time indicated that nature was more important than nurture in determining the underlying competence.

However, some studies find that nurture is also important on personal achievement. Dahl and Lochner (2005) showed that the increase of parental income through the Earned Income Tax Credit (EITC) benefits had positive effects on children's math and reading achievement. Deng and Treiman (1997) used the 1982 Chinese Census to study the impact of the Cultural Revolution on educational attainment in China. They found that the advantage of people from educated or cadre² family was drastically reduced. However, their sample was restricted to men living in multiple generation households (MGHs). Meng and Gregory (2007) used two urban surveys to study the impact of the CR on 4-year and 2-year college degree attainment. They found that the CR significantly reduced the probability of obtaining a 4-year college degree and increased the probability of obtaining a 2-year college degree. Thus, the authors argued that the 2-year college degree was a substitute for the 4-year college degree for people affected by the CR. Zhou (2014) used an urban sample from 1995 and 2002 to study the long-term impact of the CR on income. He found that the CR had lasting negative effects on permanent income for people born in 1946-1961. The negative impact of the CR on income is through educational attainment and work experience, marital instability, health, and personal attitude.

² As referenced in Deng and Treiman (1997): cadres are political appointees whose responsibilities principally include political management of the work unit to which they are attached. Paid by their work unit or firm, cadres are Communist Party functionaries in the workplace.

Recent studies also focus on the multi-generational mobility, which is the change of status across more than two generations within a family. Figuring out the pattern of the multi-generational transmission will help to further investigate the mechanism behind the intergenerational coefficient. The most widely-used empirical model assumes the intergenerational process to be an AR (1). If this assumption holds, we should expect the multi-generational correlation to die out at a geometric rate. For instance, Clark (2015) argued that the similarity of the social statuses across generations can be explained by the transmission of the underlying competence which is an AR (1) process. Thus, the multi-generational correlation is simply the intergenerational correlation to the power of the number of generations. Since he obtained large intergenerational correlations, the correlation was still large between a generation and a much later generation. However, Stuhler (2012) argued that estimating the multi-generational correlation using such a simple assumption was an “iterated regression fallacy”. Although the error term ε_{it} is constructed to be uncorrelated with parental status, it is not necessarily uncorrelated with grandparent's status. Thus, we also need to consider the direct effect of grandparent status on grandchild's status. The model in Becker and Tomes (1976) suggested that after controlling for father's earnings, the correlation between the grandfather's earnings and the grandchild's earnings was negative and small, suggesting that the direct effect of grandfather's income is insignificant. Some empirical studies on the multi-generational mobility in the US, like Lucas (2013) and Adermon et al. (2018), confirmed this inference. However, Zeng and Xie (2014) found a significantly positive effect of grandfather's schooling on grandchild's educational achievement when the grandfather co-resided with the grandchild in rural China. Solon et al. (2015) suggested that the discrepancy showed

that there is not a universal pattern for multi-generational mobility and that estimates of multi-generational mobility may suffer biases from endogeneity or measurement errors.

In this chapter, we use Chinese survey data from the CHIP (2013) which covers both urban and rural population and weight it to be geographically representative of the national level. We use the weighted data to estimate trends in intergenerational educational mobility in China in the 20th century. Then we study the impact of the Cultural Revolution on intergenerational and mutigenerational mobility, as well as the underlying mechanism linking the father's education with children's educational attainment. We contribute to the literature by applying a more representative data to study the intergenerational and multi-generational mobility in China, as well as analyze the mechanism behind the effect of the CR on the intergenerational mobility in China.

1.3 Background

1.3.1 The Intergenerational Mobility in 20th Century China

Intergenerational mobility measures inequality of opportunity across two generations. A low level of intergenerational mobility means less opportunity for people from poor families to improve their social status during their lifetime and a lower likelihood that those from rich families will move down. Variations in the level of intergenerational mobility in a region can be explained by the changes of the regional return to human capital, policy, and demographic structure (Solon, 2004). In the 20th century, China experienced tremendous institutional transitions which changed policies, return to human capital, and demographic structure. In the first two decades of the 20th century, the last feudal dynasty, the Qing, was replaced by a republican government. After the Second Sino-Japanese War and the Chinese Civil War, the People's Republic of China was established in 1949. This

was followed by several political movements, including the most radical one, the Cultural Revolution (1966-1976). After Mao's death in 1976, China returned to a more traditional merit-based educational system. The institutional changes of 20th century brought social chaos, but they also provided opportunities for people at the bottom of society.

In a paper that studied the intergenerational educational mobility in China using urban survey data, Chen et al. (2015) found an inverted U-shape pattern of the intergenerational mobility in 20th century. The level of mobility in China was low for those received schooling in the pre-Mao era (people born before the mid-1940s), reached its highest level in the Mao-era (for those born in the mid-1940s to the mid-1960s), and returned to a low level in the post-Mao era (for those born after the mid-1960s).

We replicate the analyses in Chen et al. (2015) using a nationally representative sample to see the overall pattern of the intergenerational mobility in 20th century China. Our replication uses a weighted Chinese Household Income Project sample, which is described in the data section. The sample covers both urban and rural residents and we have developed weights so that it is representative of the major regions in China, in contrast to the urban sample used in Chen et al. (2015). Rural residents make up a large proportion of the Chinese population. They accounted about 82% of the population during the CR, and still accounted for 42% of the population in 2017. Therefore, using a sample that covers the full population provides us more meaningful estimated results at the national level.

The replication results for intergenerational schooling coefficients for people born in 1920-1985 are plotted in Figure 1.1 for 5-year birth cohorts (the estimated numbers are listed in APTable 1.1). As can be seen from the graph, the estimated intergenerational

schooling coefficients exhibit a U-shaped pattern, indicating that intergenerational mobility exhibits an inverted U-shape in the 20th century. We thus confirm the pattern in Chen et al. (2015) using a sample that includes both urban and rural populations. The intergenerational schooling mobility reached its highest level for those who were born around 1960, almost all of whom experienced the Cultural Revolution (1966-1976) during their principal schooling years. We also plot the intergenerational coefficients when we recode years of education by rank and the intergenerational schooling correlations in APFigures 1.1 and 1.2 (corresponding to APTables 1.2 and 1.3). Both exhibit an inverted U-shape pattern for intergenerational educational mobility.

To check whether the observed pattern holds for different demographic groups, we estimate intergenerational schooling coefficients by current residential areas (urban or rural), which we plot in APFigure 1.3, by gender in APFigure 1.4, and by birth areas (urban or rural) in APFigure 1.5. Although different groups exhibit different levels of intergenerational coefficients, they all show an inverted U-shape pattern of intergenerational mobility.

1.3.2 Changes of the Educational System in the Cultural Revolution

Our replication of Chen et al. (2015) suggests that people received their education in the Cultural Revolution had the highest level of intergenerational educational mobility. This may relate to the severe changes of the educational system in China during the CR period.

During the Cultural Revolution, the national college entrance exam was suspended over the entire period from 1966 to 1976. At the same time, the government relaxed the

standard for people from lower family backgrounds to enter college. In the later period of the CR (1972-1976), people with lowest socioeconomic backgrounds, or those who showed their passion and loyalty to the party and government could be enrolled into college by recommendations of local leaders, without taking the entrance exams. The college entrance exams were resumed in 1977.³

Due to frequent political disruptions during the CR, schools were closed for years. In the beginning of the CR (1966-1968), almost all schools (primary, secondary, tertiary) were closed, and many teachers were forced to leave their positions. The "Down to the Countryside Movement" sent urban high school students (those who were in senior high school in 1968) to rural areas to do farm work, and they were not allowed to come back to cities until the 1970s.

The curriculum taught in schools were also changed greatly in the CR, and previous teachers were often replaced by new teachers without academic qualifications. Students were taught to do farming and manufacturing work rather than standard curricula in school. The curriculum in junior high school returned to normal in 1970, and the curriculum in senior high school returned to normal in 1972.

At the same time, primary and secondary education were greatly expanded in the rural areas of China. According to Andreas (2009), the number of senior high schools in rural areas increased from 604 to 50,916 during the CR, and the number of junior high schools

³ In the first four years after the CR, the government allowed those who had not been able to take the exams during the CR to take them regardless of their ages. This policy compensates those who delayed their entrance to college due to the CR. But from 1981 to 2001, people who were older than 25 were prohibited from taking the exams.

increased from 8,628 to 131,265. The primary and secondary enrollment rates in the rural areas increased dramatically as a result.

In summary, the Cultural Revolution suspended the college entrance exam, changed the regular curriculum, closed schools for years and sent urban students to the countryside, which greatly damaged the educational system in China. According to Du (2003), the CR caused a decrease in human capital accumulation of 14.3% based on average years of schooling. Meanwhile, the primary and secondary education were expanded in rural areas and children from lower family backgrounds were able to enter college based on recommendations from political leaders. We thus hypothesize that changes of China's educational system during the CR reduced the gap in education between children with higher socioeconomic family background and children from lower family background.

1.4 Data

In this chapter, we use the 2013 Chinese Household Income Project (CHIP) to conduct our analyses. CHIP is a cross-sectional survey that covers urban and rural respondents in 14 provinces in China in 2012. The urban and rural sample in the CHIP selects households that had formally registered residential addresses in urban or rural areas at the time of the survey⁴, respectively.

⁴ In China, the official household registration system is called "Hukou". It records the information of the registered household address and the Hukou type, either urban (non-agricultural) or rural (agricultural). The type of Hukou depends on whether the household engages in agricultural or related work. It is possible for a household to change the address or the type of Hukou, but in order to do so, household members must satisfy certain requirements established by their local governments. The Hukou system is important in China since it is directly related to welfare services a person can receive from the local government, including education, various subsidies, and pensions. In 2014, a new type of Hukou called "Resident Hukou" was created to replace the previous urban and rural Hukous. (In other words, all households have single type of Hukou). But it was initially implemented only in Henan province.

The questions in this survey covered basic demographic information of every person living in the surveyed households, such as gender, age, education, and occupation. It asked for information about the parents and children of the main respondent couples, regardless of whether the parents lived in the household, in contrast to surveys that only provide information on parents living with their children (Deng and Treiman, 1997). The detailed information on parents and children of the main respondent couples makes the 2013 CHIP a suitable sample for us to study intergenerational mobility in China across time.

To check whether the CHIP is representative of the Chinese population, we compare it with the *Chinese Statistical Yearbook 2013* (0.1% population survey statistics for the year 2012). The comparisons in APTables 1.4-1.7 suggest that the CHIP is not a nationally representative sample. The CHIP under-sampled the urban population, the population of the eastern provinces, and over-sampled the population of the central provinces (APTable 1.4). Also, it over-sampled larger-size households, 15–64-year-olds in central and western provinces and under-sampled the migrant population⁵ (APTable 1.5 and APTable 1.6). In terms of education, the CHIP over-sampled people with at least senior high school in all age groups, under-sampled people with primary or junior high school educations but over-sampled the population that were never educated among those older than 40 (APTable 1.7).

Since the original CHIP sample is not nationally representative, it is essential to re-weight the CHIP in order to improve its representation of the population. We weight it by using the fraction of “urban”, “rural”, and “migrant”⁶ residents in each province from the

⁵ In these tables, we follow official government practice of defining “migrants” as people living outside of their official registered residential area in 2013, as indicated by their Hukou membership.

⁶ The “urban” group here refers to people who lived in urban areas and had not change their residential area as of 2013. “Rural” group refers to people who lived in rural areas and had not change their residential area as of 2013. The “migrant” group refer to people who had changed their residential areas by 2013.

Chinese yearbook 2013. The details of our weighting procedure are described in Appendix section A1.

We compare the summary statistics of the weighted CHIP and the Census in APTables 1.4-1.7. After weighting, the misrepresentation problems in regional population, size of households, migrant proportions were successfully eliminated, but it helped little in correcting the over-representation problem for educated people; nor did it solve the under-representation problem for people older than 40 with less than junior high school degrees. However, these remaining differences are minor.

The descriptive statistics of respondents and their spouses who were at least 25 years old in the 2013 CHIP (after weighting) are listed in Table 1.1. There are 32,976 individuals in our sample whose ages were older than 25. 28% of the weighted CHIP sample had obtained at least a high school degree and 11% had obtained at least a college degree. The average years of schooling, high school degree attainment rate, college degree attainment rate, average years of father's schooling, and average years of mother's schooling all increase over cohorts. The growth rates of parents' schooling across cohorts are larger than the growth rates for children.

We also divide the sample into three categories by whether the years individuals would have attended primary or higher education were covered by the regime of Mao: pre-Mao (for those born before 1945), Maoist (for those born in 1945-1963), and post-Mao (born after 1963). It suggests that the average years of schooling increased by 17% from the pre-Mao cohort to the Mao cohort, but it increased at a much higher rate (30%) from the Mao cohort to the post-Mao cohort. These statistics indicate that the human capital accumulation

rate in the Mao cohort was slower than in the post-Mao period, which is likely related to the institutional changes in the Mao era. The decrease in growth of schooling years differs by family background, as can be seen from Figure 1.2. The average years of schooling in the CR increased least—and perhaps declined—for people whose fathers had at least a high school education, and the gaps of average years of schooling by father’s education shrank greatly in the CR. Average education for those whose fathers attended primary school increased in all three periods, but growth was particularly fast for the Mao cohort.

1.5 Methods

In this chapter, we use the difference-in-difference method to identify the impacts of the Cultural Revolution on intergenerational and multi-generational mobility. The extent to which schooling for each individual is affected by the CR is measured by indexes that are based on a respondent’s birth year. We compare the cohort differences (measured by the CR indexes) to the differences in educational attainment for people with different parental backgrounds.

People who received their education during the CR were not affected in a uniform manner. For instance, those who were in the last year of high school in 1966 may have been affected by the CR because they lost the chance to take the college entrance exam at the normal time; but people who were in middle schools in 1966 were also affected by the closure of schools, changes in curriculum, and the “Down to the Countryside” movement. We assume the impact of the CR differed by the number of years during school age a person was affected. Meng (2007) lists a measure of the number of affected years for people pursuing education in the CR period (see Column 3 in Table 1.2) bases on birth year and interruptions in education. The “affected years” in this measure include both the years of

school closure and the years in which schools did not follow a standard curriculum. Alternatively, Zhou (2014) measured the affected years in a different way (see Column 4 in Table 1.2). His version of “affected years” includes only the years of school closure. Both measures of the CR-affected years display an inverted U-shape pattern during 1966-1976, and zero before and after.

Using these measures of affected years, we first model the impact of the CR on the intergenerational educational mobility using a linear probability model:

$$EDU_{i,t} = \sum_{j=1}^3 \beta_j (Ftype_{i,j} * YA_i) + \beta_4 YA_i + \sum_{j=1}^3 \beta_{j+4} Ftype_{i,j} + \gamma \mathbf{X}_i + \delta_t + \varepsilon_{i,t} \quad (1.1)$$

$EDU_{i,t}$ is a dummy variable indicating whether a person obtained at least a high school degree ($HS_{i,t}$), or a college degree ($CD_{i,t}$) in cohort t , YA_i is the index for the number of years a person i was affected by the radical educational interventions during the CR, using either Meng's or Zhou's measures (see Table 1.2). YA_i equals 0 for those who would have received their education in periods other than the CR. $Ftype_{i,j}$ is father's education which is categorized into four levels ($j = 1,2,3,4$, correspondingly): at least high school degree, middle school degree, primary school degree and less than primary school education (the base group). \mathbf{X}_i are control variables including cohort-variant effects of gender, age, age squared, father's age and father's age squared, people's Hukou type at birth (agricultural or non-agricultural). δ_t is child's birth cohort fixed effects, which is a dummy for those born in the pre-CR cohorts. β_1 , β_2 , and β_3 measure the impact of duration of the CR for people with three different levels of father's education, relative to those whose fathers had less than a primary education. β_4 measures the impact of the CR on education for people whose

fathers with less than primary education. β_5 , β_6 and β_7 measure the effects of the levels of a father's education on children's educational attainment for children whose schooling were not affected by CR. These three coefficients are of interest to us since they estimate the DiD effect of the CR on the relative advantage of having a father with different levels of education.

We also study impacts of the CR on intergenerational mobility for people grew up in rural and urban areas.

$$\begin{aligned}
EDU_{i,t} = & \beta_1 YA_i + \beta_2 Rural_i + \sum_{k=1}^3 \beta_{3k} Ftype_{i,k} + \beta_4 (YA_i * Rural_i) \\
& + \sum_{k=1}^3 \beta_{5k} (YA_i * Ftype_{i,k}) + \sum_{k=1}^3 \beta_{6k} (Ftype_{i,k} * Rural_i) \quad (1.2) \\
& + \sum_{k=1}^3 \beta_{7k} (YA_i * Ftype_{i,k} * Rural_i) + \gamma \mathbf{X}_i + \delta_t + \varepsilon_{i,t}
\end{aligned}$$

$Rural_i$ is a dummy for people who were born in rural areas. Since rural-urban migration was extremely low in the Maoist period (1949-1976), we assume people grew up and received most of their education in the same type of areas where they were born. A partial exception is the “Down to countryside movement”, which increased urban-rural migration for senior high school students, discontinuing their education at the time. YA_i is the index for the years affected by the CR using Meng's measure. All the other variables are the same as those in equation (1.1). β_4 estimates the difference in the impact of the CR for those who grew up in rural areas relative to those who grew up in urban areas and whose father had

only a primary school education; $\beta_{6,1}$, $\beta_{6,2}$, $\beta_{6,3}$ estimate the differences in effect of parental education for rural-born people, relative to urban-born people. $\beta_{7,1}$, $\beta_{7,2}$, $\beta_{7,3}$ estimate the differences in impact of the CR by parental education for people born in rural areas, relative to people born in urban areas.

We then study the impact of the CR on multi-generational educational transmission. Specifically, we study whether the transmission pattern between grandfather and grandchild changed if the father was affected by the CR:

$$EDU_{i,t} = \beta_1(GFS_i * YA_i) + \beta_2 YA_i + \beta_3 GFS_i + \gamma \mathbf{X}_i + \delta_t + \varepsilon_{i,t} \quad (1.3)$$

$$EDU_{i,t} = \alpha_1(FS_i * YA_i) + \alpha_2 YA_i + \beta_3 FS_i + \gamma \mathbf{X}_i + \delta_t + \varepsilon_{i,t} \quad (1.4)$$

$$EDU_{i,t} = \omega_1(GFS_i * YA_i) + \omega_2(FS_i * YA_i) + \omega_3 YA_i + \omega_4 GFS_i + \omega_5 FS_i + \gamma \mathbf{X}_i + \delta_t + \varepsilon_{i,t} \quad (1.5)$$

$EDU_{i,t}$ is a dummy for whether a grandchild (child of the main respondent) obtained at least a high school degree ($HS_{i,t}$), or a college degree ($COL_{i,t}$) in cohort t . GFS_i and FS_i are the years of schooling for the grandfather (father of the main respondent) and the father (the main respondent), respectively. YA_i is the index for the years the father (the main respondent) was affected by the CR using Meng's measure. Control variable \mathbf{X}_i includes grandchild's gender, age, age squared, number of siblings in equations (1.3)-(1.5); grandfather's age, grandfather's age squared in equations (1.3) and (1.5); and father's age and its square in equations (1.4) and (1.5). δ_t denotes the birth cohort fixed effects for grandchild (dummies of birth cohorts in before 1962, 1962-1967, 1968-1972, 1973-1977, 1978-1982). The estimate of β_1 shows us whether the relationship between a grandfather's

schooling and grandchild's schooling was weakened by the CR either directly or indirectly. Equation (1.4) estimates whether the CR has any effect on the next generation. If having a father who is affected by the CR reduces the advantage of having a more educated father for the next generation, we would expect to find negative estimate of α_1 in equation (1.4). Finally, we include both the grandfather's and the father's years of schooling in the model in equation (1.5), which estimates how the CR altered the direct effect of the grandfather's education on the grandchild's education through ω_1 .

Next, we study the mechanism behind the effect of father's education on the child's college degree attainment by decomposing the process into three parts: obtaining a high school degree, participating the college entrance exam (CEE), and obtaining a CEE score:

$$COL_{i,t} = \beta_1(FS_i * YA_i) + \beta_2YA_i + \beta_3FS_i + \alpha A_i + \gamma X_i + \delta_t + \varepsilon_{i,t} \quad (1.6)$$

A_i denotes the step indicators of college degree attainment. We have four model specifications: the base specification with no step indicator, a specification with high school degree indicator, a specification with both high school degree and CEE participation indicators, and a full specification with three of the indicators (high school attainment, CEE participation, and CEE scores). All specifications control for other demographic variables (age, age squared, gender, father's age and father's age squared, Hukou status at birth), cohort fixed effects and residential province fixed effects of children. The full specification includes the step indicator variables, HS_i (high school degree attainment), CEE_i (whether person i took college entrance exam), and $CEEscore_i$ (the CEE scores divided by 750, the highest possible score). To allow comparability across different model specifications in equation (1.6), we assign the CEE scores for those who did not report their scores the

sample average score and create a dummy for this missing group. We also assign the year of taking the CEE for those who did not report it to be the year when they were 19 years old and create a dummy for those people. In addition, we add a dummy for people who did not report the province where they took the CEE and a dummy for the type of CEE exam (science and engineering, social science and history, and art and PE) taken. All the other variables in equation (1.6) have the same definition as shown in equation (1.2).

Comparing the estimated coefficients of β_1 across model specifications allows us to calculate the contribution of each of the step indicators to the effect of the CR on child's educational achievement by father's schooling. We can also calculate the contribution of each of the step indicators on the impact of father's schooling on child's education for those not affected by the CR by comparing the changes in estimated coefficients of β_3 .

Finally, we use occupation as a proxy for father's status, to examine whether the impact of the CR on the intergenerational mobility changes using different status proxy. Also, we study the effects of mother's schooling on child's educational attainment and the impact of the CR on this relationship using:

$$EDU_{i,t} = \beta_1(MS_i * YA_i) + \beta_2YA_i + \beta_3MS_i + \gamma X_i + \delta_t + \varepsilon_{i,t} \quad (1.7)$$

$$EDU_{i,t} = \alpha_1(FS_i * YA_i) + \alpha_2YA_i + \beta_3FS_i + \gamma X_i + \delta_t + \varepsilon_{i,t} \quad (1.8)$$

$$EDU_{i,t} = \omega_1(MS_i * YA_i) + \omega_2(FS_i * YA_i) + \omega_3YA_i + \omega_4MS_i + \omega_5FS_i + \gamma X_i \quad (1.9)$$

$$+ \delta_t + \varepsilon_{i,t}$$

where MS_i denotes years of the mother's schooling, and FS_i denotes years of the father's schooling. All the other variables in equations (1.7)-(1.9) have the same definition as

shown in equation (1.2). The control variable X_i includes the mother's age and its squared in equations (1.7) and (1.9), and it also includes father's age and age squared in equations (1.8) and (1.9). In equation (1.9) ω_1 estimates how the CR changed the direct effect of mother's schooling on child's educational attainment and ω_4 estimates the direct effect of mother's schooling on child's educational attainment in the non-CR period.

1.6 Results

1.6.1 The Impact of the CR on Intergenerational Educational Mobility

1.6.1.1 General Impact of the CR on Intergenerational Educational Mobility

The debate over the importance of genetic factors and environmental factors on intergenerational transmission never stops (Mazumder, 2008, Björklund et al., 2005, Clark, 2015, Dahl, 2012). Analyzing the impact of the Cultural Revolution on the intergenerational mobility provides new empirical evidence for this topic since the CR is a radical environmental change for Chinese students in the period 1966-1976.

Table 1.3 lists the estimated coefficients for the effect of the CR on intergenerational mobility, as shown in equation (1.1). Columns (1) and (2) report the results on high school diploma attainment, columns (3) and (4) report the results on the college degree attainment, and columns (5) and (6) report the results on whether a person obtained a college degree and enrolled in the college before age 20. We use Meng's measure of years affected by the CR in columns (1), (3) and (5), and Zhou's measure in columns (2), (4) and (6).

Table 1.3 suggests that for those not affected by the CR, both sets of results imply that having a father with a high school degree (or above) increases the likelihood of the child obtaining a high school degree by about 40 percentage points and increases the likelihood

of obtaining a college degree by about 31 percentage points, relative to those whose father with less than primary school education, as can be seen from the coefficients of Ftype1 from the table.

The table also suggests that the advantages of having a high school degree (or above) decreased significantly for a child experiencing more years of the CR educational disruption. Each year affected by the CR decreased the advantage of having a father with a high school degree (or above) by about 1.8 percentage points in terms of high school attainment and college degree attainment when using Meng's measure (column (1), Ftype1*YA). Therefore, for those affected by the CR, on average the advantage of having fathers with high school degrees declines by about 5.8 percentage points (the average years affected by CR is 3.2), which shrinks the advantage by about 14.6%⁷ compared with those born in non-CR cohorts when Meng's measure is used. For college degree attainment, the decrease of the advantage of having a father with high school degree is on average about 18.5% for people whose fathers with at least high school degree when Meng's measure is used. The estimated decreases in the advantage of having a father with a junior high school or primary high school education on either obtaining a high school degree or a college degree due to the CR are negative but small and statistically insignificant. People whose fathers had less than primary school education were 0.8 percentage points less likely to obtain a high school degree per year affected by the CR, relative to those who received their educations in other periods. This reflects the negative impact of the CR on human capital. If we restrict college degree attainment for those who obtained a college degree

⁷ This was obtained by using the average decrease rate in the advantage, 5.8, divided by the advantage of having a high school degree fathers in other cohorts, 39.5.

and enrolled in college before age 20⁸, the negative effect of the CR interruptions on reducing the advantage of having a high-educated father on college degree attainment is even greater, as can be seen in column (5).

Using Zhou's index, which measures only the years affected by closing schools in the CR, we find similar pattern as observed using Meng's measure, though the magnitude of the estimated coefficients are slightly smaller. The discrepancy between the results using Zhou's measure and Meng's measure suggests that school closure alone has smaller impact on changing the association between father's education and children's education than the impact when combining with the changes in curricula.

Although the interruptions in education during the CR successfully reduced the advantage of having an educated father, the advantage was not eliminated by the CR. For example, even when a person experienced the maximum possible disruption during the CR (8 years), having a father in the highest education category still increased the chance of receiving a high school degree by 25 percentage points, relative to those whose fathers have less than primary school education, 16 percentage points higher than those with fathers with a primary school education, and 7 points higher than those whose fathers with a middle school education. Similarly, for people affected by the CR for the maximum possible years, those with a high school educated father were still 17 percentage points more likely than those with a father had less than primary school education to obtain a college degree, 14 percentage points more likely than those with a primary school educated

⁸ After the CR, Chinese government allowed people to take the College Entrance Exam (CEE) without any age restriction until 1981. In 1982-2001, the CEE only allowed people younger than 25 years old. Therefore, allowing people to take the CEE can be seen as a compensation policy for the CEE suspension in the CR.

father, and 8 percentage points more likely than those with a father with a middle school education.

Some may argue that people's residential province is an important confounder since it may affect both father's education and children's education. However, in the CHIP survey, there is no data for the province where a person was born or grew up. Instead, we have data for the residential province at the time of the survey, which may be a function of education, as individuals often migrate after obtaining education. In APTable 1.8, we include a residential province fixed effects to address the potential endogeneity problem. The patterns observed in APTable 1.8 are similar to Table 1.3, but the estimated effects are smaller than those in Table 1.3. Given the possibility that including province fixed effects may induce bias, we have chosen to focus on results that omit such controls.

To check the robustness of the above results, we drop older respondents from the sample, measure the CR interruptions as schooling missing patterns, use a sample of siblings and control for family fixed effects, apply Logit and Probit models, and change the school entry age in Meng's measure.

In particular, one possible source of bias is that older respondents may be a selected sample since they survived and were able to take the survey themselves, indicating that they are more advantaged relative to others in the same cohort. Thus, we omit people who were older than 67 years of age from the sample, i.e., we omit the pre-CR group, and re-estimate equation (1.1), leaving about 94% of the full sample. The estimated results can be seen in the APTable 1.9, which are similar to our estimates in Table 1.3.

The results report the marginal effects based on the Logit and Probit regressions are shown in APTable 1.10. The estimated results for high school attainment are similar with the results in Table 1.3, but the estimated coefficients on the advantage of having a father with a high school degree or higher on college degree attainment are smaller in magnitude than in those in Table 1.3. This may be due to the small sample for respondents with college degrees. Nevertheless, the basic patterns we observe in APTable 1.10 are similar to those of the linear probability results.

Some may argue that the interruptions in the CR are not linearly related to people's educational achievement. We use schooling missing pattern dummies according to Table 1.2 (including missing high school education only, missing high school and middle school, missing high school middle school and primary school, missing middle school and primary school, and missing primary school education only), and their interaction terms with years of father's schooling, to substitute the "YA" measure and its interaction terms. The results are reported in APTable 1.11. We can observe that each schooling missing pattern in the CR decreased the advantage of having a more educated father in obtaining a high school or a college degree, which is consistent with the findings in Table 1.3.

We also use information on sibling education that is obtained from respondents. Although this approach increases the effective sample size, it may also introduce systematic errors if education of siblings is misreported. We have weighted cases by the inverse of the total number of siblings in a family, so the distribution of family background will not change. We find that when replicating our analysis using the same methods as described above, coefficient estimates for interactions between the CR and father's educational type (reported in APTable 1.12) are somewhat larger, although patterns are

unchanged. We control for family fixed effects to estimate the same model to see if it is the differences between families that causes the patterns we observed in Table 1.3. The results are shown in the APTable 1.13, which show similar patterns to those shown in Table 1.3 and confirms that the CR decreased the advantage of having an educated father. This suggests that the negative effect of the CR on the advantage of having an educated father in educational achievement also occurred within a family.

Meng's measure assumes the school entry age to be 7, which is the minimum age of school entry. We change the school entry age to 6 and 8 in Meng's measure and see if that affects our results. These results for Meng's measure of YA are listed in APTable 1.14, which are similar to the results in Table 1.3.

1.6.1.2 Area Differential Impacts of the CR on the Intergenerational Educational Mobility

During the Cultural Revolution, primary and secondary schools were greatly expanded in rural areas. We explore whether this expansion changed the general pattern of the CR on intergenerational mobility we observed in Table 1.3 by estimating the coefficients in equation (1.2). The estimated results are listed in Table 1.4. Column (1) lists the estimated results for high school degree attainment, and column (2) lists the results for college degree attainment.

The differences in the effects of the CR between urban and rural areas on the intergenerational associations are significant, as can be seen from the first three rows in Table 1.4. The coefficients on three-way interaction terms indicate that the negative effect of the CR index on obtaining a high school degree for those who have a father with a high school degree or above is 7.8 percentage points greater in absolute value for people grew

up in the rural areas than in the urban areas. Also, the negative effect of the CR index on obtaining a college degree for those who have a father with a middle school diploma is 1.4 percentage points greater in absolute value in the rural areas than in the urban areas. For the effect of the CR on those whose fathers had less than primary education, the negative effect is much smaller in the rural areas than in the urban area (see the coefficients of “YA*Rural”). For people grew up in urban areas, each year the CR significantly decreased the advantage in terms of obtaining a college degree for those whose fathers had a high school education or above ($Ftype1*YA$), those whose fathers had a middle school education ($Ftype2*YA$), and whose fathers had a primary school education ($Ftype3*YA$); however, the negative effect of the CR on people who grew up in urban areas on obtaining a high school diploma is not statistically significant.

The coefficient of the interaction term between the CR index and the rural dummy indicates that the negative effect of the CR on those whose fathers had less than a primary education is much smaller in the rural areas than in the urban areas. People living in rural areas whose fathers had less than a primary school diploma are 1.7 percentage points more likely to obtain a high school diploma and 1 percentage point more likely to obtain a college degree for each year that the CR affects their schooling relative to comparable individuals in urban areas. The CR increased the probability of obtaining a college degree for people from rural areas whose fathers had less than a primary school education. Each year affected by the CR increased the probability of obtaining a college degree for people who grew up in rural areas by 0.4 percentage points. We also find that for those who are not affected by the CR, the probability of obtaining a high school degree or a college degree is much

greater for people living in urban areas than those living in rural areas (see the interaction of father's education and the rural dummy).

These results suggest that for those not affected by the CR there is a gap in educational attainment between urban residents and rural residents, and the gap is greater as father's education decreases. The CR significantly expanded the urban-rural gap for people whose fathers had at least a high school degree but reduced the gap for people from the least educated family. The CR benefited people from the least advantaged families from rural areas in obtaining a college degree.

1.6.2 The Impact of the CR on Multi-Generational Mobility

Multi-generational mobility is the variation of status across more than two generations within a family. Since the most widely used empirical model assumes the intergenerational process to be an AR (1) process, if this assumption holds, we expect the multi-generational correlation to die out at a geometric rate. Many studies provided empirical evidence to support this assumption, showing that the direct correlation between grandfather's social status and grandchild's social status is weak (Ridge, 1973, Behrman, 1985, Warren, 1997, Lucas, 2013). However, Stuhler (2012) argues that estimating the multi-generational correlation using such a simple assumption is an "iterated regression fallacy". Grandparent's status may also have an important direct impact on grandchild's status. Zeng et al. (2014) found that the grandfather's schooling had a significantly positive effect on a grandchild's educational achievement when grandparents are co-resident with their grandchildren in rural China. Solon (2018) suggested that the discrepancies in the literature imply that there is not a universal pattern for multi-generational mobility.

We estimate the direct and indirect effects of the grandfather's schooling on grandchild's education, as well the impact of the CR on these effects in equations (1.3)-(1.5). One of the issues we face is that the CHIP does not have complete, three-generation information, as it lacks information on children living outside of the household for rural respondents. Complete information is only available for urban respondents. However, if we only use data on urban residents, it may cause estimation bias. Luckily, what we found in the APFigure 1.3 is that the patterns of intergenerational mobility are similar for the urban sample and rural sample. Thus, we have at least some confidences that the results for the multi-generational relationship using the urban sample will be similar to that for the rural sample. We also compare the intergenerational estimated results using the full sample to the results using only the urban sample and the urban sample of families with children in APTable 1.15. Based on this comparison, we expect the results to be similar using the urban sample to those using full sample if it were available.

The estimated results for equations (1.3)-(1.5) using the urban sample are listed in columns (1), (2), (3) for high school degree and columns (4), (5), (6) for college degree attainment in Table 1.5. First, we see that each additional year of the grandfather's schooling increases the probability of obtaining a high school degree by 2 percentage points and increases the probability of obtaining a college degree by 2.3 percentage points (GFS). Adding father's schooling into the model in columns (3) and (6) suggests that the direct effect of the grandfather's schooling on grandchild's educational attainment is significantly positive. Comparing columns (1) and (3), (4) and (6), we see that about 10% of the effect of grandfather's schooling on child's high school degree attainment is transmitted from the effect of father's schooling, and father's schooling accounts for about 17% of grandfather's

schooling on grandchild's college degree attainment. The result is consistent with the finding in Zeng et al. (2014) that the direct effect of father's schooling on grandchildren's education is significant in China. We also find that the CR had insignificant impact on the relationship between grandfather's schooling and grandchild's educational achievement, which is consistent with Clark (2015) that the long-run association across generations within a family cannot be significantly altered by radical political events or movement. Clark (2015) obtained his conclusion by tracing the average status within families with rare surnames for several centuries in various countries, while our results are obtained by analyzing the impact of the CR using regular DiD method. Second, the CR did not significantly change the association of education between grandfather and grandchild. In other words, grandchildren of the bad class and the upper-middle class were helped by better-educated grandparents than grandchildren of the worker and peasant classes.

We also estimate the effect of experiencing the CR for the father's generation on the child's educational attainment in column (2) and column (5). Interestingly, the results suggest that each year of disruption due to the CR would increase the probability of obtaining a college degree for his child by about 0.24 percentage points. It is likely that those who lost opportunities to receive education tended to invest more in their children's education, which is consistent with the findings in Roland and Yang (2017). They may treat such investment as a compensation to their own loss from the CR. The results are also consistent with those of Xie and Zhang (2019), although their approach uses a different indicator for CR impact and codes education differently.

To summarize this subsection, we find that grandfather's schooling and father's schooling significantly and positively influence children's education, and that these effects

are mostly direct. Also, we do not find evidence for declines in the ability of grandfather or fathers to pass on the educational advantage to their grandchildren or children when the father's generation was affected by the CR. In contrast, the advantage from having a more educated father in attaining a college degree is even stronger for people who had fathers with schooling affected by the CR.

1.6.3 Potential Mechanisms

We have shown that the CR reduced the educational advantage of having a more educated father for those facing CR disruptions during their principal school ages. But we are still not clear about the mechanism explaining such reduction of advantage. We study this question using equation (1.6) to decompose the impacts of father's schooling on child's college degree attainment. It also allows us to identify how the CR affects the intergenerational transmission through each of the step factors.

To obtain a college degree (either three-year or four-year) in China, a person must first obtain a high school diploma. If the person pursues a four-year college degree, he/she needs to take the national college entrance exam (CEE)⁹, and the college acceptance depends on CEE scores.¹⁰ Thus, several key issues jointly determine whether a person can obtain a college degree: obtaining a high school degree, taking the CEE, obtaining a sufficient CEE scores to meet the admission requirement of a specific college, and finishing all college course requirements to obtain the college degree. Parental background can directly affect

⁹ The exam was administered annually beginning in 1952 but was suspended in 1966-1976. The exam can be taken more than once but in 1981-2001, it only opened to high school graduates who were younger than 25.

¹⁰ A small number of students are enrolled into college through other channels. For example, students who win awards in math, physics, chemistry or biology contests and students admitted for outstanding athletic or artistic accomplishments.

the probability of children obtaining a college degree attainment, as well as indirectly affecting the probability through affecting each step in the process. When a radical political intervention occurs, like the Cultural Revolution, it may weaken the direct or the indirect links between a person's parental background and their college degree attainment.

In the CHIP sample, there are 32,981 father-respondent pairs. Among them, 9,182 respondents are at least high school graduates, which is about 28% of the sample. 3,364 respondents reported that they had taken the CEE, which is 11% of the full sample and 37% of all high school graduates. 2,697 people reported their CEE scores, which amounts to 8% of the full sample and 80% of people who reported they took the CEE. Finally, 3,620 people had college degrees (three-year and above) in the sample, which is about 134% of people who reported their CEE scores, 108% of people who reported they took the CEE, 39% of all high school graduates and 11% of the full sample. The number of college graduates exceeds the numbers of reported CEE takers, which may occur because some three-year colleges did not require the CEE score for admission, and some respondents who took the CEE may not have reported it. One of the potential biases in conducting this analysis is that the CEE participation rate and scores in our sample may not be representative. In Appendix A2, we compare the CEE participation rate using CHIP data with the rate obtained from the Census 2000 and an administrative online file¹¹, and do the comparison in APFigure 1.6. These comparisons suggest that CEE participation in the CHIP is similar to that of the full population.

¹¹ “Number of People Took CEE and College Acceptance Rate in 1949-2012”, Baidu Wenku (in Chinese). UCL: <https://wenku.baidu.com/view/e4a5434b2b160b4e777fcf04.html>, last visit: May 8, 2018.

The estimated results of the four model specifications of equation (1.6) are listed in Table 1.6. The specifications assume a sequential order structure for the measures that contribute to obtain a college degree, and the differences between the estimated coefficients provides the contribution of each step indicator to the mechanism. We observe that the coefficient of father's schooling for those who were not affected by the CR is decreasing as we add more step indicators into the model. For instance, when there is no step indicator, it suggests that an additional year increase in the father's schooling improves the probability of obtaining a college degree by 2 percentage points. When adding all of three indicators into the model, the estimated coefficient drops to 0.6 percentage points. The result suggests that these three step indicators jointly explain about 70% of the association between father's schooling and the college degree attainment of children. Specifically, high school completion explains about 50%¹² of the association, taking the CEE explains about 15%, and obtaining a higher CEE scores explains another 5%, leaving about 30% of the coefficient unexplained by these three step indicators.

In addition, we observe the mechanism of how the advantage of having a more educated father was reduced by the CR. When no step indicator was included, the reduction of advantage per year of father's schooling is 0.14 percentage points per year affected by the CR, and when we include all of the step indicators in the model, the reduction of the advantage drops to 0.02 percentage points per year affected by the CR, suggesting that the CR reduced the advantage of having a father with more schooling through these three step indicators by about 86%. Specifically, obtaining a high school degree explains about 36%

¹² This was calculated using the coefficient of the father's schooling in model (2) minus the coefficient of father's schooling in model (1) and divided by the coefficient of the father's schooling in model (1).

of the reduction in the advantage, attending the CEE explains about 29%, and obtaining a high CEE score explains about 21%. We use a bootstrap method to obtain the standard error of the differences for the estimated coefficients of the interaction term between father's schooling and YA.

We also use the decomposition method in Gelbach (2016) in the APTable 1.16 to check how does the three step indicators contribute to the procedure. We may note that Gelbach's decomposition method does not consider the sequential order of adding explanatory variables to the regression. The specifications listed in Table 1.6 do follow a sequential order and thus it is more suitable for our research purpose.

1.7 Extensions

1.7.1 The Intergenerational Association between Parental Occupation and Children's Education

To this point, we have measured the intergenerational relationship between parental education and children' education. One of the unique features of the CR is that people were classified by their parental occupation (see footnote 1), which largely determined how government and other people treated them in the CR period. The normally "advantaged" types of parental occupation became "inferior" ones during the CR, and that may have made it difficult for people who were in the in school during the CR to obtain more education. Previous studies suggest that the CR decreased the advantage of the privileged classes but benefited those from the bottom classes in terms of education achievement (Deng & Treiman, 1997; Giles et al, 2019; Xie & Zhang, 2019). This subsection applies the CR index and the DiD structure regression to estimate the impact of father's and grandfather's occupations on child's educational attainment. We re-estimate equations

(1.1), (1.3), (1.4), (1.5), and change the main explanatory variables to be dummies for four types of parental occupations: 1. Principal, manager, technician or professional; 2. Clerk, commercial and service personnel; 3. Farmers or manufacturing related personnel; 4. Other types of occupations. We also include their interactions with the CR.

Estimate of the association between child's educational attainment and father's occupation, and between child's educational attainment and grandfather's occupation are shown in APTables 1.17 and 1.18. In APTable 1.17, we find a pattern similar to that of the educational measure in Table 1.3, that having a higher-status occupation father helps in obtaining high school and college degrees. Also, when children were affected by the CR, the advantage of having a father with a higher-status occupation on the educational attainment was significantly reduced. A difference between the results in APTable 1.17 and Table 1.3 is that we observe a significant decline in the advantage of having a father who was in the second and the third occupational categories. This may suggest that the policies related to family backgrounds in the CR were more closely tied to parental occupation than parental educational level.

In APTable 1.18, we study the impact of grandfather's occupation on child's educational attainment. Having a grandfather who held a higher-status occupation helps a grandchild in obtaining both high school and college degrees. However, we find that having a grandfather who worked as clerk or commercial personnel helped a grandchild to obtain a high school or a college degree than those with a manager or technician grandfather. But the advantage of having a clerk or commercial personnel grandfather was significantly decreased if the father's generation was affected by the CR. This result is different from our conclusion using education as proxy for grandfather's status, suggesting that the

transmission of grandfather's advantage to grandchild was significantly reduced for people with a father who worked as a clerk or in commercial activities.

1.7.2 The Intergenerational Educational Association between Mother's Schooling and Child's Education

Some studies, such as Heckman et al. (1986), Schultz (1994) and Thomas et al. (1996) have found that mother's schooling is more important for children than the father's schooling. Other studies, like Behrman and Rosenzweig (2002), suggest that the observed association between the mother's schooling and the children' achievement may be due to heritability and assortative mating. In this section, we use the CHIP sample to analyze the impact of the mother's schooling on high school and college degree attainment, the effects of the CR, and the role of assortative mating in explaining the relationship in equations (1.7)-(1.9).

The estimation results using equations (1.7), (1.8) and (1.9) are reported in APTable 1.19. We observe similar effects of mother's schooling and father's schooling on high school and college degree attainment. When we control for both father's and mother's education, we see that each of their coefficients declines about 40%. Hence the importance of assortative mating is similar in explaining observed effects for fathers and mothers. However, the decrease of the advantage of having a more educated mother on college degree attainment in the CR is mainly because more educated mothers are more likely to have more educated fathers—and it is father's education that experiences a penalty during the CR. The different effects of the CR for the mother's and father's schooling are likely related to policies during the CR, which directly targeted people with higher status fathers rather than higher status mothers.

1.8 Conclusion

In this chapter, we study the impact of the interruptions during the Cultural Revolution on intergenerational educational mobility, multi-generational mobility and analyze the mechanisms behind it. Using survey data that covers both urban and rural population in China and weighting it to be geographically representative at the national level, we obtain more reliable estimates than previous studies. We confirm the results in Chen et al. (2015) that the intergenerational educational mobility exhibits an inverted U-shape pattern in 20th Century China, with the highest mobility occurring for those who received education in the years of radical change in the educational system.

We study the impact of the CR on high school and college degree attainment, and how the impact differed across levels of parental education. We find that the advantage of having a high school educated father on the probability of obtaining a high school or college degree was reduced by about 2 percentage points for every additional year the child's schooling was affected by the CR. The decrease in the advantage of having a father with at least a high school degree is greater for people who grew up in rural areas. The negative impact of the CR on educational achievement is smaller for people from traditionally disadvantaged family background who grew up in rural areas, due to the school expansion in rural areas. Although the CR decreased the advantage of having an educated father on child's educational attainment, the advantage of having an educated father was not eliminated.

We decompose the process of obtaining a college degree to that of obtaining a high school degree, participating in the College Entrance Exam (CEE), and obtaining a high CEE score. We find that the advantage of having a more educated father helps a person in

obtaining a college degree mainly through each of these steps. The interruptions during the CR also reduced the advantage of having a more educated father through these three steps.

We also study the impact of the CR on the multi-generational transmission of educational achievement. Grandfather's schooling has a direct positive impact on grandchild's educational attainment. Having a father who received his education in the CR did not reduce the advantage of having an educated grandfather on grandchild's education. Interestingly, we find that the longer a father received education in the CR, the higher is the probability of his children obtaining a college degree. The impact of the CR on intergenerational and multi-generational mobility provides empirical evidence on the debate of nature versus nurture in explaining the association between father's education and children's education. The CR interruptions cut off certain channels for parents to nurture their children, and thus reduced the association between father's education and children's education, suggesting the importance of nurture in the transmission. However, in the long run, the CR interruptions did not reduce the association between grandfather's education and grandchild's education. We may infer that stable family characteristics are important in explaining the transmission in the long run.

The CR benefited people whose fathers worked as farmers or manufacturing workers in obtaining college degrees but reduced the advantage of having a father with traditional "better occupations". We also estimate the effect of maternal schooling on children's educational attainments. Assortative mating explains about a third of the observed effect of either paternal or maternal education on the child's educational attainment. We do not observe any significant decline in the advantage of having an educated mother for those affected by the CR after controlling for father's schooling.

The Cultural Revolution was marked as a radical change in China's educational system. It shut down many educational resources and decreased the quality of education, but it also provided new educational opportunities to rural children and children from traditionally disadvantaged family backgrounds. Previous to the CR, most people in China were illiterate, as only those from higher socioeconomic family backgrounds could afford education. Therefore, the destruction of the previous educational system in the CR had large negative impacts on those from what had been more privileged family backgrounds but had little impact for most others. Meanwhile, the policies in the CR benefited people in rural areas and those from poor family backgrounds, increasing their access to education, providing a level of opportunity that was nearly impossible before the CR.

Table 1.1 Descriptive Statistics of the Weighted CHIP Sample

	Mean	Std	Min	Max	N
Panel A: Pre-Mao Cohort (Born before 1945)					
Age	75.09	4.59	70	98	1,915
Male	0.58	0.49	0	1	1,915
Years of Schooling	6.39	4.28	0	19	1,915
High School Degree Attainment	0.17	0.38	0	1	1,915
College Degree Attainment	0.06	0.24	0	1	1,915
Years of Father's Schooling	2.05	3.30	0	19	1,915
Years of Mother' Schooling	0.77	2.07	0	15	1,915
Panel B: Maoist Cohort (Born before 1945-1963)					
Age	58.45	5.20	51	69	13,568
Male	0.52	0.50	0	1	13,568
Years of Schooling	7.47	3.59	0	20	13,568
High School Degree Attainment	0.23	0.42	0	1	13,568
College Degree Attainment	0.06	0.23	0	1	13,568
Years of Father's Schooling	2.86	3.69	0	19	13,568
Years of Mother' Schooling	1.43	2.82	0	19	13,568
Panel C: Post Mao Cohort (Born after 1963)					
Age	41.16	6.21	25	50	17,493
Male	0.49	0.50	0	1	17,493
Years of Schooling	9.74	3.53	0	21	17,493
High School Degree Attainment	0.33	0.47	0	1	17,493
College Degree Attainment	0.16	0.36	0	1	17,493
Years of Father's Schooling	5.74	4.23	0	19	17,493
Years of Mother' Schooling	4.09	4.01	0	19	17,493
Panel D: All Cohorts					
Age	49.98	11.77	25	98	32,981
Male	0.50	0.50	0	1	32,981
Years of Schooling	8.66	3.78	0	21	32,981
High School Degree Attainment	0.28	0.45	0	1	32,981
College Degree Attainment	0.11	0.31	0	1	32,981
Years of Father's Schooling	4.40	4.21	0	19	32,981
Years of Mother' Schooling	2.86	3.69	0	19	32,981

Table 1.2 Measures of Affected Years in the Cultural Revolution

Birth Year	Age in 1966	CR Affecting Yrs.- Meng and Gregory (2007)	CR Affecting Yrs.- Zhou (2014)	Primary and Secondary Schooling Missing Patterns (Meng et al, 2007)	Collee Entrance Delay Patterns
(1)	(2)	(3)	(4)	(5)	(6)
1947	19	0	0.5	--	
1948	18	1	1	Missing Senior	
1949	17	2	1	High School Only	College Closed
1950	16	3	4	(1948-1950)	
1951	15	4	4	Missing Junior and	(1947-1954)
1952	14	5	6	Senior High School	
1953	13	6	6	(1951-1953)	
1954	12	7	6	Missing Primary, Junior and Senior High School	
1955	11	8	4		Recommendati on Only
1956	10	5	4	Missing Primary and	
1957	9	4	1	Junior High School	(1955-1958)
1958	8	3	1	(1956-1958)	
1959	7	3	0.5	Missing Primary	No Entrance Delay but Faced Intense Competition
1960	6	2	0	School Only	
1961	5	1	0	(1959-1961)	
1962	4	0	0	--	(1959-1963)
1963	3	0	0		

Note: This table shows the schooling miss patterns during the CR and measures of "years affected by the CR on education" in Meng and Gregory (2007) and Zhou (2014). Column (1) shows each of the birth years for people in their prime schooling years during the CR; column (2) lists the age of those people at the beginning of the CR (1966); column (3) lists the "affected years" in the CR for those people using Meng's measure based on birth year; column (4) lists the "affected years" in the CR using Zhou's measure based on birth year; columns (5) and (6) list possible schooling miss patterns in the CR using Meng's measure, the birth years of the affected population in each pattern are listed in parentheses. Column (5) shows the miss patterns of primary and secondary school, and column (6) lists the possible delay patterns of college entrance, with affected birth year listed in parentheses.

Table 1.3 Estimated Coefficients of the Impact of Affected Years in the CR on Education

	High School Degree		College Degree		College Degree (Enrollment Age <=19 yrs old)	
	(1)	(2)	(3)	(4)	(5)	(6)
Ftype1*YA	-0.018** (0.007)	-0.015* (0.008)	-0.018* (0.009)	-0.013 (0.011)	-0.025*** (0.006)	-0.031*** (0.008)
Ftype2*YA	-0.001 (0.004)	0.001 (0.008)	0.003 (0.005)	-0.003 (0.007)	-0.011*** (0.001)	-0.013*** (0.001)
Ftype3*YA	-0.003 (0.002)	-0.005 (0.003)	-0.002 (0.002)	-0.003 (0.003)	-0.002* (0.001)	-0.003*** (0.001)
YA	-0.0077* (0.0039)	-0.007** (0.002)	-0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.002* (0.001)
Ftype1	0.395*** (0.014)	0.390*** (0.015)	0.313*** (0.024)	0.310*** (0.024)	0.176*** (0.018)	0.173*** (0.017)
Ftype2	0.252*** (0.018)	0.251*** (0.018)	0.137*** (0.017)	0.140*** (0.017)	0.057*** (0.009)	0.056*** (0.009)
Ftype3	0.115*** (0.014)	0.115*** (0.014)	0.042*** (0.008)	0.043*** (0.007)	0.013** (0.005)	0.013** (0.005)
Measure	Meng	Zhou	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.3007	0.3003	0.2446	0.2441	0.1680	0.1665
N	32,981	32,981	32,981	32,981	32,981	32,981

Note: This table reports the regression results for obtaining high school and college degrees based on equation (1.2). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Columns (1) and (2) report the results on high school diploma attainment, columns (3) and (4) report the results on the college degree attainment, and columns (5) and (6) report the results on whether a person obtained a college degree and enrolled into the college before age 20. Ftype1 denotes a person's father with high school degree; Ftype2 denotes a person's father has middle school diploma; Ftype3 denotes a person's father has primary school diploma. YA is the measure of years of schooling affected by the CR, using either Meng's measure or Zhou's measure. Control variables include cohort-variant demographic variables (gender, Hukou's type at birth, age and age squared, father's age and father's age squared) and child's birth cohort fixed effects. The standard errors were clustered at the child's residential province level.

Table 1.4 Estimated Coefficients of the Impact of the CR on Education by Area of Birth

	High School Degree (1)	College Degree (2)
Ftype1*YA*Rural	-0.078*** (0.016)	-0.013 (0.012)
Ftype2*YA*Rural	-0.006 (0.018)	-0.014* (0.008)
Ftype3*YA*Rural	0.008 (0.005)	0.001 (0.004)
Ftype1*YA	-0.002 (0.010)	-0.037*** (0.008)
Ftype2*YA	-0.001 (0.011)	-0.013*** (0.003)
Ftype3*YA	-0.009* (0.004)	-0.002* (0.001)
YA*Rural	0.017*** (0.005)	0.010** (0.004)
YA	-0.027** (0.008)	-0.006*** (0.002)
Ftype1*Rural	-0.031 (0.029)	-0.214*** (0.025)
Ftype1	0.441*** (0.024)	0.440*** (0.028)
Ftype2*Rural	-0.143*** (0.028)	-0.206*** (0.020)
Ftype2	0.347*** (0.016)	0.279*** (0.025)
Ftype3*Rural	-0.105*** (0.028)	-0.088*** (0.017)
Ftype3	0.192*** (0.022)	0.111*** (0.038)
Rural	-0.302*** (0.017)	-0.110*** (0.011)
Cohort-variant Demog. Var.	Yes	Yes
Cohort FE	Yes	Yes
R ²	0.3049	0.2520
N	32,981	32,981

Note: This table reports the regression results of equation (1.4). Column (1) lists the estimated results for high school degree attainment, and column (2) lists the results for college degree attainment. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Rural is a dummy for whether a person's Hukou at birth is the agricultural type. YA uses Meng's measure of the CR interruptions. All of the other variables in the table are the same as shown in Table 1.3. Control variables include cohort-variant demographic variables (gender, age and age squared, father's age and father's age squared) and child's birth cohort fixed effects. The standard errors were clustered at the child's residential province level.

Table 1.5 Estimated Coefficients for the CR Impacts on Multi-Generational Educational Transmission

	High School Degree			College Degree		
	(1)	(2)	(3)	(4)	(5)	(6)
GFS*YA	-0.002 (0.001)		-0.002 (0.002)	-0.001 (0.002)		-0.001 (0.002)
FS*YA		0.001 (0.002)	0.001 (0.001)		0.0024** (0.0009)	0.0026** (0.010)
YA	0.024 (0.031)	0.007 (0.018)	0.036 (0.039)	0.008 (0.033)	-0.008 (0.009)	0.003 (0.030)
GFS	0.020** (0.007)		0.018** (0.006)	0.023** (0.010)		0.019* (0.009)
FS		0.037*** (0.006)	0.036*** (0.006)		0.044*** (0.004)	0.042*** (0.004)
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
R ²	4,680	4,680	4,680	4,680	4,680	4,680
N	0.080	0.1657	0.1701	0.1165	0.2183	0.2241

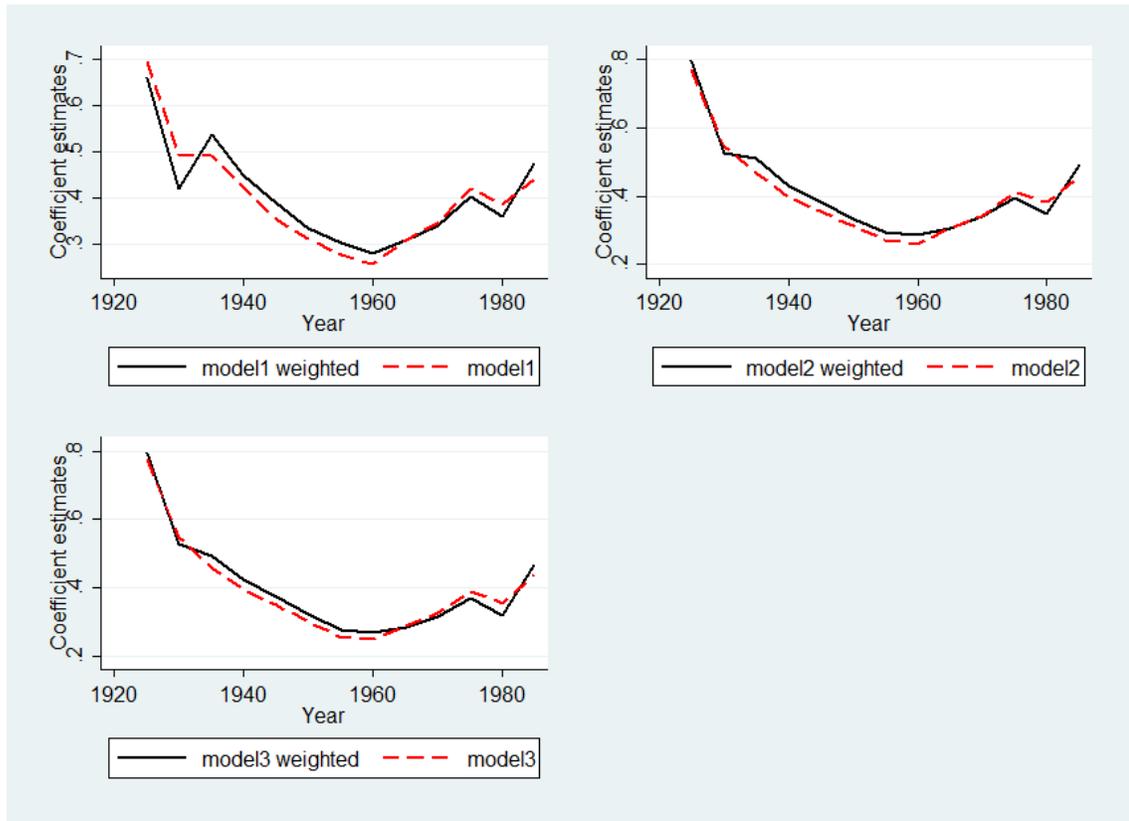
Note: This table reports the regression results of equation (1.3) (corresponding to columns (1) and (4)), equation (1.4) (corresponding to columns (2) and (5)), and equation (1.5) (corresponding to columns (3) and (6)) in terms of high school degree and college degree attainment of grandchildren. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. YA denotes years father was affected by the CR; FS denotes years of father's schooling; GFS denotes years of grandfather's schooling; In column (1), we control the cohort-variant grandchild's gender, age, age squared, grandfather's age, grandfather's age squared, Hukou's type at birth; in column (2), we control the cohort-variant grandchild's gender, age, age squared, father's age, father's age squared Hukou's type at birth; in column (3), we include all controls in (1) and (2). In all of the specifications, grandchild's birth cohort fixed effects are controlled. The standard errors were clustered at the grandchild's residential province level.

Table 1.6 Decomposition of the Effect of Father's Schooling on College Degree Attainment

	No Controls	Add HS	Add CEE Participation	Add CEE Score
	(1)	(2)	(3)	(4)
FS*YA	-0.0014*** (0.0005)	-0.0009** (0.0004)	-0.0005 (0.0004)	-0.0002 (0.0003)
YA	0.003 (0.003)	0.005 (0.004)	0.005 (0.003)	0.004 (0.003)
FS	0.020*** (0.002)	0.010*** (0.001)	0.007*** (0.001)	0.006*** (0.001)
HS	--	0.349*** (0.019)	0.209*** (0.020)	0.220*** (0.021)
CEE	--	--	0.419*** (0.021)	0.175 (0.106)
CEE score	--	--	--	1.111*** (0.001)
Explained FS*YA	--	-0.0005*** (0.0001)	-0.0004*** (0.0001)	-0.0003*** (0.0001)
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R ²	0.2340	0.3860	0.4934	0.5307
N	32,981	32,981	32,981	32,981

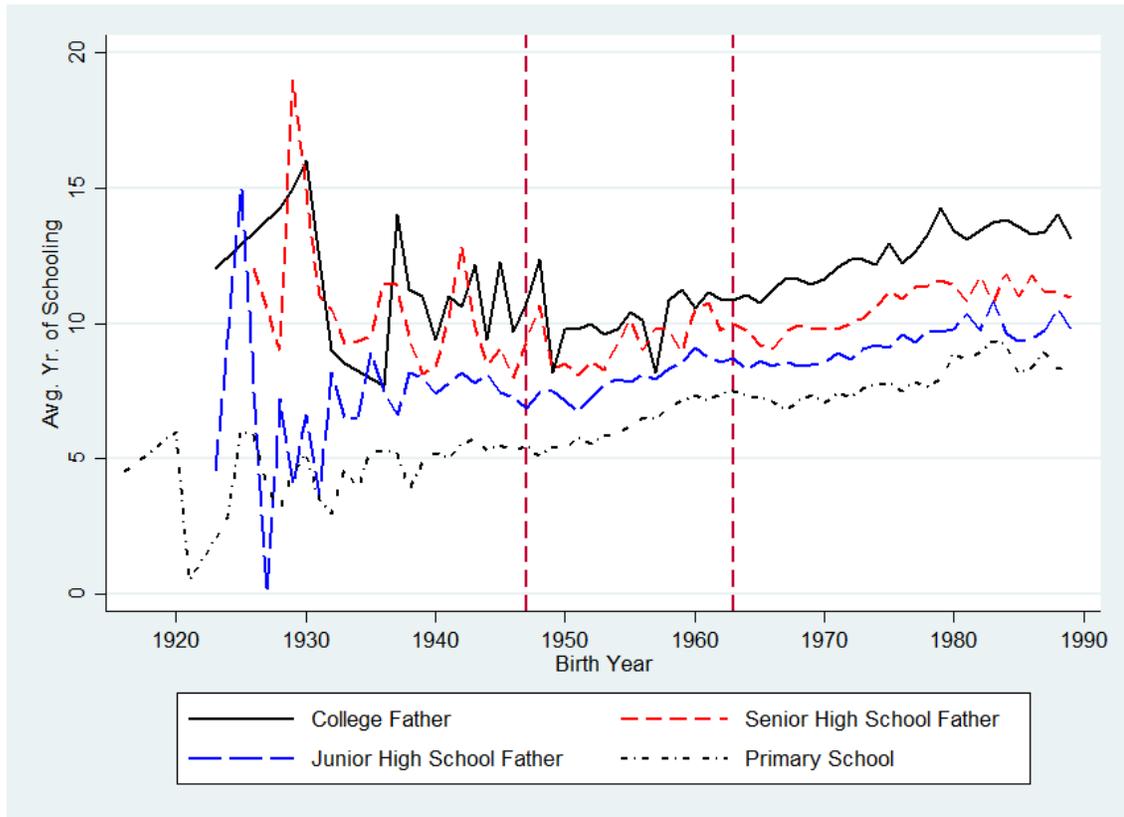
Note: The left panel in the table reports the regression results for the base specification and the full specification using equation (1.6). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The base specification does not include HS (whether the person obtained a high school degree), CEE (whether the person took the CEE), and CEEscore (the CEE score divided by 750, the highest possible score), but it includes other cohort-variant demographic variables (age, age squared, gender, father's age and father's age squared, Hukou status at birth), cohort fixed effects. The full model specification includes step variables (HS, CEE, CEEscore) and other relevant controls for the CEE scores (the year of taking the CEE, the province of taking the CEE, the CEE exam types and the missing value dummies for these variables). The standard errors were clustered at the child's residential province level. The "Explained FS*YA" row calculate the difference of the estimated coefficients from the corresponding column to the previous column, SE are obtained using the bootstrap method.

Figure 1.1 Estimated Intergenerational Educational Coefficients in 1925-1985



Note: These graphs were plotted based on the estimated intergenerational schooling coefficients for the three model specifications in Chen et al. (2015). The regression model in that paper is: $EDU_{i,t} = \sum_{t=1}^{13} \beta_t FS_{i,t} + \gamma \mathbf{X}_{i,t} + \delta_t + \varepsilon_{i,t}$. $EDU_{i,t}$ is years of schooling of a person i born in cohort t , which is one of the five-year cohorts from 1925 to 1989 (1925-1929, ..., 1985-1989, 13 cohorts in total). $FS_{i,t}$ is the years of father's schooling. δ_t is a full set of 5-year birth cohort dummies from 1925 to 1989. In model specification 1 (upper left), $\mathbf{X}_{i,t}$ is the empty set; in model specification 2 (upper right), $\mathbf{X}_{i,t}$ contains cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared; in the model specification 3 (lower left), $\mathbf{X}_{i,t}$ contains all controls in model 2, and it also includes the cohort-specific effects of living in a coastal province. The solid lines are estimated intergenerational coefficients weighted by the weights described in appendix section A1, and the dash lines are estimated coefficients without weighting.

Figure 1.2 Average Years of Schooling across Birth Year by Father's Education



Note: The figure plots the average years of schooling across birth year and by father's educational level. The solid line denotes the average years of schooling for people whose father with college degree, the short dashed line denotes the average years of schooling for people whose father with a high school degree, the long dashed line denotes the average years of schooling for people whose father with a junior high school degree, and the dotted line denotes the average years of schooling for people whose father with primary school education (Source: Chinese Household Income Project 2013).

1.9 Appendix

A1. CHIP Data Weighting Procedure

To improve the representativeness of the CHIP, we weight it by based on the fraction of urban, rural, and migrant¹³ residents in each region from the Chinese yearbook 2013. Our weighting process is summarized as follow: first, we categorize the overall population into 12 strata: metropolitan-urban, metropolitan-rural, metropolitan-migrant, eastern-urban, eastern-rural, eastern-migrant, central-urban, central-rural, central-migrant, western-urban, western-rural, western-migrant.¹⁴ Second, we weight each stratum based on the corresponding population in each stratum. Since there are only 14 provinces in the CHIP, we also weight each province so that the 14 provinces can be jointly representative for their corresponding regions. Thus, the weighting process is divided into two parts: in the first part, it corrects the sample in each stratum of each province; in the second part, it further corrects the sample to be representative at the region level. Thus, the weight for a person i living in a province j (a province in the 14 provinces in the CHIP) in the stratum k can be written as:

$$w_i^{j,k} = \frac{N_j^k}{n_j^k} \times \frac{N^k}{\sum_j N_j^k}$$

Where N_j^k is the population in province j of stratum k in the *Yearbook*, n_j^k is the CHIP sample size in province j of stratum k , N^k is the *Yearbook* population of stratum k , and $\sum_j N_j^k$ is the sum of *Yearbook* population in provinces belonging to a stratum k in the

¹³ The migrant group here refers to people who did not live at their official registered address.

¹⁴ The eastern provinces include Hebei, Liaoning, Jilin, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong; the central provinces include Shanxi, Anhui, Jiangxi, Hubei, Henan, Hunan; western provinces include Inner Mongolia, Guangxi, Sichuan, Yunnan, Guizhou, Xizang, Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang; and metropolitan areas include Beijing, Shanghai and Chongqing.

CHIP. For example, Liaoning, Jiangsu, Shandong and Guangdong are the four eastern provinces in the CHIP whereas people from other eastern provinces such as Zhejiang and Fujian were not included in the CHIP. Thus, the population of the eastern-urban stratum N^k includes all the urban population of the eastern provinces of China, while $\sum_j N_j^k$ includes only the urban population in Liaoning, Jiangsu, Shandong and Guangdong since they are the only eastern provinces in the CHIP.

A2. Summary Statistics for CEE Participation and CEE Scores in CHIP

The CHIP survey asked respondents and their household members whether they had ever taken CEE. If they had, they were also asked what year they took the CEE and what their CEE scores were. However, a large fraction of sample did not answer this question (59% of the sample). Moreover, we are not sure whether the CEE participation rate calculated in our sample can truly represent the actual participation rate. Therefore, we compare some descriptive statistics on CEE participation from the sample with statistics calculated from administrative data and the Chinese census and see if the CHIP is representative in terms of CEE participation.

The first step is to check the reason for the large missing proportion of respondents for the question on whether or not they took the CEE. First, we list the frequency and percentage of the sample who answer “yes”, “no”, or those who did not answer this question by their birth cohort. We can see in APTable 1.20 that the missing group is the largest group in each cohort. The missing portion is about 60% of the whole sample. We then restrict the sample to those who had at least 12 years of schooling (at least finishing high school). We find that the missing portion dramatically decreases to below 10%, indicating that most of the respondents who left the question blank did not have high school

diplomas. Since a person needs to finish high school courses in order to take the CEE, those who did not answer this question were not able to take CEE. In panel (3), we compare several demographic statistics for people with at least a high school degree. We can see that the average age, average schooling, father's schooling and mother's schooling for the missing group is similar to those who answered "no". They are older, have less schooling, have less father's schooling and less mother's schooling than the "yes" group. This further confirms that the missing group is very likely to be people who did not take the CEE. Therefore, in our later analysis, we recode this missing group as "no".

Next, we compare the CEE participation rate of the 19-year-old population between the CHIP and administrative statistics. Here, we assume most people took the CEE at age 19 (the exact age does not matter if we use the same standard to compare the CHIP and the census). Therefore, the CEE participation rate for "the CHIP sample is calculated by using the number of people who report they took the CEE in the CHIP sample in a certain year divided by the number of 19-year-olds in the CHIP in that year. Similarly, we calculated the CEE participation rate from the census using the number of people who took the CEE in a certain year based on administrative data divided by the estimated number of 19-year-olds in the population in the same year. Here, we use the information from the Chinese census 2000 and an administrative online file from Baidu Wenku, and do the comparison in APFigure 1.6.¹⁵

¹⁵ "Table 3-1 Number of Population by Age Group and Gender" "Table 6-1 Number of Deaths by Province and Age Group and Gender (1999.11.1-2000.10.31)" (The Fifth Chinese Census, 2010). "Number of People Took CEE and College Acceptance Rate in 1949-2012", Baidu Wenku (in Chinese). UCL: <https://wenku.baidu.com/view/e4a5434b2b160b4e777fcf04.html>, last visit: May 8, 2018.

In APFigure 1.6 we observe that the percentage of people who took the CEE in the CHIP has a similar pattern to that calculated from the census: the rate is small (<6%) before the restart of the CEE in 1977; and it surges in 1977-1981, during which people delayed by the CR were allowed to take the CEE at any age. Beginning in 1981, only those younger than 25 were allowed to take CEE, so the rate drops dramatically in that year. After 1981, the rate increases over time. Although the trend calculated from the CHIP under-represents the CEE rate in 1977-1979, 2003-2008 and over-represents the rate in 1993-2000, the general pattern is very similar to that of the census. The over-representation of the rate in the CHIP in these periods could be related to the over-representation of more educated respondents in the CHIP relative to the population, but we do not have a good explanation for the under-representation.

We also calculate the percentage of the CEE takers by their birth year using the CHIP (the horizontal axis represents the year when they were 19 years old). This series (the dotted line in APFigure 1.6) shows that the percentage of people taking the CEE increased for those who were born after 1954 (aged 19 in 1973). We observed that the number of people who were born in 1945-1957 (aged 19 in 1964-1976) and took the CEE is greater than number of people who took the test during 1964-1976, suggesting that most people born in 1945-1957 took the exam after the CR (since the exam was suspended in 1966-1976), most likely in 1977-1981, when no age restriction applied. Also, the number of people who took the exam in 1977-1981 is much greater than the number of people who were 19 years old in that period and took the test in any period, indicating most CEE takers in this period were from previous cohorts (older than 19 years old). After 1985, the two trends become

very similar, suggesting that the majority of CEE takers are those that are 19-year-olds after 1985.

Since the statistics of the CEE participation in the CHIP is similar to the population, we can thus use the sample to do the analysis on the mechanism linking parental schooling with children's college degree attainment.

APTable 1.1 Estimated Intergenerational Educational Coefficients by Birth Cohort, Replicating Chen et al. (2015): Years of Schooling

Years of Schooling	(1)	(2)	(3)
1925*FS	0.660*** (0.089)	0.795*** (0.115)	0.795*** (0.120)
1930*FS	0.422*** (0.080)	0.526*** (0.071)	0.523*** (0.075)
1935*FS	0.537*** (0.056)	0.510*** (0.058)	0.494*** (0.057)
1940*FS	0.449*** (0.037)	0.431*** (0.038)	0.422*** (0.037)
1945*FS	0.389*** (0.028)	0.382*** (0.028)	0.373*** (0.029)
1950*FS	0.334*** (0.020)	0.335*** (0.020)	0.322*** (0.020)
1955*FS	0.303*** (0.019)	0.296*** (0.019)	0.275*** (0.019)
1960*FS	0.282*** (0.014)	0.285*** (0.014)	0.270*** (0.013)
1965*FS	0.309*** (0.013)	0.307*** (0.014)	0.285*** (0.013)
1970*FS	0.341*** (0.016)	0.340*** (0.016)	0.316*** (0.016)
1975*FS	0.404*** (0.021)	0.393*** (0.022)	0.369*** (0.020)
1980*FS	0.361*** (0.029)	0.349*** (0.030)	0.319*** (0.030)
1985*FS	0.474*** (0.045)	0.490*** (0.050)	0.466*** (0.050)
Cohort FE	Yes	Yes	Yes
Baseline Control	No	Yes	Yes
Extended Controls	No	No	Yes
N	32,976	32,976	32,976
R ²	0.2867	0.3192	0.3387

Note: The table reports the estimated intergenerational coefficients based on Chen et al. (2015): $EDU_{i,t} = \sum_{t=1}^{13} \beta_t FS_i + \gamma X_i + \delta_t + \varepsilon_{i,t}$. $EDU_{i,t}$ is years of schooling of a person i born in cohort t , which is one of the five-year cohorts from 1925 to 1989 (1925-1929, ..., 1985-1989, 13 cohorts in total). In the table, FS_i measures the father's years of schooling. X_i includes cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared. In the extensive model's specification, it also includes the cohort-specific effects of living in a coastal province. t is a full set of 5-year birth cohort dummies from 1925 to 1989. Standard errors are clustered at child's residential province level.

APTable 1.2 Estimated Intergenerational Educational Coefficients by Birth Cohort, Replicating Chen et al. (2015): Rank of Schooling

Years of Schooling	(1)	(2)	(3)
1925*FSR	0.570*** (0.113)	0.680*** (0.110)	0.678*** (0.113)
1930*FSR	0.410*** (0.075)	0.497*** (0.070)	0.498*** (0.075)
1935*FSR	0.456*** (0.056)	0.425*** (0.058)	0.494*** (0.075)
1940*FSR	0.434*** (0.038)	0.414*** (0.039)	0.402*** (0.038)
1945*FSR	0.414*** (0.030)	0.408*** (0.030)	0.396*** (0.030)
1950*FSR	0.378*** (0.022)	0.377*** (0.022)	0.361*** (0.022)
1955*FSR	0.350*** (0.022)	0.337*** (0.022)	0.310*** (0.022)
1960*FSR	0.365*** (0.018)	0.375*** (0.018)	0.351*** (0.018)
1965*FSR	0.395*** (0.017)	0.394*** (0.018)	0.362*** (0.018)
1970*FSR	0.408*** (0.019)	0.406*** (0.020)	0.375*** (0.019)
1975*FSR	0.452*** (0.024)	0.436*** (0.025)	0.409*** (0.024)
1980*FSR	0.382*** (0.029)	0.370*** (0.030)	0.337*** (0.030)
1985*FSR	0.501*** (0.050)	0.509*** (0.053)	0.484*** (0.050)
Cohort FE	Yes	Yes	Yes
Baseline Control	No	Yes	Yes
Extended Controls	No	No	Yes
N	32,976	32,976	32,976
R ²	0.1523	0.1870	0.2108

Note: The table reports the estimated intergenerational coefficients based on Chen et al. (2015): $EDU_{i,t} = \sum_{t=1}^{13} \beta_t FSR_i + \gamma X_i + \delta_t + \varepsilon_{i,t}$. $EDU_{i,t}$ is rank of schooling of a person i born in cohort t , which is one of the five-year cohorts from 1925 to 1989 (1925-1929, ..., 1985-1989, 13 cohorts in total). The rank is the relative position of a father's schooling among all fathers in the same birth cohort t , and it ranges between 0 to 1 (1 indicates the highest possible father's schooling in that cohort). In the table, FSR_i the rank of father's schooling. X_i includes cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared. In the extensive model's specification, it also includes the cohort-specific effects of living in a coastal province. t is a full set of 5-year birth cohort dummies from 1925 to 1989. Standard errors are clustered at child's residential province level.

APTable 1.3 Estimated Intergenerational Educational Correlations by Birth Cohort, Replicating Chen et al. (2015): Years of Schooling

Years of Schooling	(1)	(2)	(3)
1925*FS	0.519*** (0.055)	0.626*** (0.073)	0.626*** (0.075)
1930*FS	0.288*** (0.037)	0.360*** (0.033)	0.361*** (0.035)
1935*FS	0.412*** (0.033)	0.391*** (0.034)	0.378*** (0.034)
1940*FS	0.369*** (0.025)	0.355*** (0.026)	0.348*** (0.025)
1945*FS	0.361*** (0.024)	0.354*** (0.025)	0.346*** (0.025)
1950*FS	0.330*** (0.019)	0.331*** (0.019)	0.318*** (0.019)
1955*FS	0.314*** (0.020)	0.306*** (0.021)	0.340*** (0.020)
1960*FS	0.355*** (0.022)	0.360*** (0.022)	0.340*** (0.021)
1965*FS	0.384*** (0.021)	0.383*** (0.021)	0.354*** (0.021)
1970*FS	0.407*** (0.023)	0.405*** (0.023)	0.377*** (0.023)
1975*FS	0.442*** (0.025)	0.430*** (0.026)	0.404*** (0.024)
1980*FS	0.384*** (0.032)	0.371*** (0.034)	0.339*** (0.034)
1985*FS	0.492*** (0.048)	0.508*** (0.053)	0.483*** (0.054)
Cohort FE	Yes	Yes	Yes
Baseline Control	No	Yes	Yes
Extended Controls	No	No	Yes
N	32,976	32,976	32,976
R ²	0.2867	0.3192	0.3387

Note: The table reports the estimated intergenerational correlations ($\rho = \beta \frac{\sigma_1}{\sigma_0}$, where β is the estimated intergenerational coefficient, and σ_0, σ_1 are standard deviations of father's schooling, child's schooling, in each of birth cohort t , respectively), based on Chen et al. (2015). $EDU_{i,t}$ is years of schooling of a person i born in cohort t , which is one of the five-year cohorts from 1925 to 1989 (1925-1929, ..., 1985-1989, 13 cohorts in total). In the table, FSR_i measures the rank of father's schooling. X_i includes cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared. In the extensive model's specification, it also includes the cohort-specific effects of living in a coastal province. t is a full set of 5-year birth cohort dummies from 1925 to 1989. Standard errors are clustered at child's residential province level.

APTable 1.4 Population Distribution by Region in the CHIP Sample and the Chinese Yearbook 2013

	CHIP	YB13	WCHIP-YB13
Panel A: Total Sample			
Metropolitan	5.1	3.3	0.0
Eastern	31.2	43.0	0.0
Central	37.7	26.7	0.0
Western	25.9	27.0	0.0
Panel B: Urban Sample			
Metropolitan	9.6	5.4	0.0
Eastern	31.5	47.8	0.0
Central	34.6	23.8	0.0
Western	24.3	22.9	0.0
Panel C: Rural Sample			
Metropolitan	3.1	0.9	0.0
Eastern	31.0	37.2	0.0
Central	39.4	30.0	0.0
Western	26.5	31.9	0.0
Panel D: Urbanization Rate			
Metropolitan	58.2	88.1	0.0
Eastern	31.5	60.4	0.0
Central	28.5	48.5	0.0
Western	29.4	46.0	0.0

Note: Panel A shows the regional distribution of the CHIP sample (column (1)), the distribution in the Chinese Yearbook 2013 (column (2)), and the differences in regional distribution between the weighted CHIP sample and the yearbook (column (3)). Panel B shows regional distribution for urban residents in the CHIP sample and the Chinese Yearbook, and the difference between weighted CHIP and the yearbook; Panel C shows the regional distribution for rural residents in the CHIP sample and the Chinese Yearbook, and the difference between weighted CHIP and the yearbook; Panel D shows the urbanization rates in each region in the CHIP and the yearbook, and the difference between the weighted CHIP and the yearbook. Column (3) is 0 by construction of the weights.

APTable 1.5 Household Distribution and Average Household Size of the 14 CHIP Provinces in CHIP Sample and the Chinese Yearbook 2013

Province	% of Households			Avg. HH. Size			Diff in HH. Size			
	CHIP (1)	YB13 (2)	WCHIP- YB13	CHIP (4)	YB13 (5)	WCHIP- YB13	1 per (7)	2 pers (8)	3 pers (9)	4 pers (10)
Panel A: Eastern Provinces										
Beijing	6.4	2.7	2.3	3.1	2.6	0.4	-11.6	3.9	11.2	-3.6
Jiangsu	8.3	9.5	0.5	4.0	2.9	1.0	-8.9	-4.4	4.0	9.3
Shandong	8.5	12.4	1.9	3.5	2.8	0.6	-8.5	-5.4	7.4	6.6
Guangdong	8.6	11.7	0.7	4.7	3.2	1.1	-14.9	-11.5	4.5	21.9
Liaoning	6.2	5.7	1.4	3.2	2.7	0.4	-6.6	5.4	0.2	1.0
Panel B: Central Provinces										
Shanxi	7.1	4.3	-1.2	3.6	3.0	0.6	-5.7	-1.5	4.1	3.0
Anhui	6.7	7.0	-2.0	4.0	3.0	0.9	-7.3	-8.7	3.8	12.1
Henan	9.0	10.2	-2.7	4.2	3.3	0.8	-6.8	-7.0	2.2	11.6
Hubei	7.2	6.9	-2.0	4.1	2.9	1.0	-6.0	-9.8	0.9	14.9
Hunan	7.1	7.3	-1.8	4.1	3.3	0.7	-10.7	-1.7	5.4	7.0
Panel C: Western Provinces										
Chongqing	5.5	3.9	0.4	3.8	2.7	1.0	-15.9	-5.7	9.3	12.3
Sichuan	7.6	10.5	1.5	3.8	2.8	1.0	-13.3	-5.2	5.8	12.6
Yunnan	6.2	5.1	0.9	4.3	3.3	0.8	-9.6	-5.3	3.3	11.7
Gansu	5.6	2.7	0.4	4.5	3.4	0.9	-7.7	-9.6	0.0	17.2

Note: Columns (1) to (3) show the proportions of household in each province in the CHIP, in the yearbook, and the difference between weighted CHIP and the yearbook; Columns (4) to (6) show the average household size in each province of the CHIP sample, in the yearbook, and the difference between weighted CHIP and the yearbook; Columns (7) to (10) show the differences of the sample distribution by household size (1 person, 2 persons, 3 persons, 4+ persons, respectively) between the CHIP and the yearbook.

APTable 1.6 Demographic Comparison between CHIP Sample and Chinese Population by Province

Province	% of Male			% of Same Hukou			Diff in HH. Size		
	CHIP (1)	YB13 (2)	WCHIP- YB13	CHIP (4)	YB13 (5)	WCHIP- YB13	CHIP (7)	YB13 (8)	WCHIP- YB13
Panel A: Eastern Provinces									
Beijing	48.9	51.8	-5.0	79.3	45.6	3.0	80.5	81.5	-1.5
Jiangsu	51.3	50.0	-1.3	91.4	75.9	0.8	74.4	74.5	0.0
Shandong	50.7	50.9	-0.4	92.0	84.7	0.9	76.9	73.6	3.4
Guangdong	52.1	52.7	-1.6	88.6	63.1	1.4	74.1	76.0	-2.0
Liaoning	50.1	50.5	-1.3	85.1	78.1	0.9	75.3	79.4	-4.3
Panel B: Central Provinces									
Shanxi	52.0	51.4	0.7	82.3	89.1	-4.1	77.5	76.3	1.3
Anhui	52.4	51.0	-0.2	93.1	87.6	2.4	74.7	71.0	4.0
Henan	52.2	50.3	1.7	92.2	94.4	-0.6	74.2	70.3	3.9
Hubei	51.6	50.8	0.5	88.6	83.5	2.1	74.2	75.2	-0.3
Hunan	51.7	51.4	0.8	90.6	84.5	6.8	74.4	71.1	3.5
Panel C: Western Provinces									
Chongqing	50.6	51.0	-0.9	87.8	78.9	4.0	73.4	71.2	2.5
Sichuan	51.4	51.0	-0.5	92.6	87.6	-0.5	73.7	70.7	3.8
Yunnan	50.8	51.8	-1.2	93.4	86.3	1.6	72.2	72.0	1.2
Gansu	52.2	51.4	0.6	95.3	82.6	6.1	76.3	74.2	3.4

Note: Columns (1) to (3) show the fraction of male in each province of the CHIP, the fraction in the yearbook, and the difference between the weighted CHIP and the yearbook statistics; Columns (4) to (6) show the fraction of non-migrant sample (Same Hukou is defined as that the registered resident address for a person is in the same county as the address he/she currently lives in. This is the common way to classify the migration status for a person in China.) in each province of the CHIP, the fraction in the yearbook, and the difference between weighted CHIP and the yearbook; Columns (7) to (9) show the fraction of 15-64 year-olds in each province of the CHIP, the fraction in the yearbook, and the difference between weighted CHIP and the yearbook

APTable 1.7 Demographic Comparison between CHIP Sample and Chinese Population by Province

Age Group	Educational Level						
	None	Primary	Junior HS	Senior HS	Associate	Undergraduate	Graduate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
25-29	-0.2	-1.7	-8.7	3.7	2.8	3.6	0.4
	(-0.3)	(-3.0)	(-12.7)	(4.4)	(3.8)	(6.9)	(0.8)
30-34	-0.2	-1.9	-7.0	2.6	2.2	3.8	0.5
	(-0.3)	(-3.1)	(-12.4)	(4.4)	(4.2)	(6.3)	(0.8)
35-39	0.1	-1.5	-8.1	3.9	0.9	3.9	0.9
	(-0.3)	(-4.4)	(-12.5)	(6.3)	(3.1)	(6.7)	(1.1)
40-44	1.2	-1.0	-5.9	2.3	1.1	1.9	0.3
	(0.5)	(-4.1)	(-8.5)	(4.7)	(2.7)	(4.0)	(0.1)
45-49	1.2	0.7	-5.3	1.4	0.7	1.4	0.1
	(0.3)	(-1.8)	(-6.8)	(3.6)	(2.2)	(2.5)	(0.1)
50-54	1.5	-4.2	-1.4	2.6	0.9	0.6	0.1
	(0.6)	(-6.3)	(-2.4)	(4.5)	(2.1)	(1.5)	(0.1)
55-59	4.0	-8.9	-1.2	6.0	-0.1	0.1	0.1
	(2.8)	(-10.2)	(-1.3)	(7.3)	(0.9)	(0.5)	(0.1)
60-64	5.4	-8.4	-1.0	3.1	0.6	0.1	0.1
	(4.1)	(-10.8)	(-0.2)	(5.0)	(1.5)	(0.3)	(0.1)
65-69	5.0	-6.1	-1.4	1.6	0.3	0.6	0.0
	(3.2)	(-7.9)	(-1.2)	(3.9)	(0.8)	(1.2)	(0.0)
70-74	2.7	-8.6	0.9	3.9	0.3	0.9	0.0
	(-1.5)	(-10.3)	(2.9)	(6.5)	(0.9)	(1.6)	(0.0)
75-79	4.8	-11.3	1.7	3.5	0.3	1.0	0.0
	(3.0)	(-13.9)	(3.1)	(5.3)	(0.8)	(1.7)	(0.0)
80-84	11.7	-15.9	1.5	1.7	-0.1	1.0	0.1
	(10.0)	(-15.9)		(2.1)	(-0.2)	(1.4)	(0.1)
>=85	14.4	-15.8	-2.1	1.9	0.6	0.6	0.3
	(11.0)	(-15.0)	(-2.2)	(4.1)	(0.9)	(0.9)	(0.2)

Note: The table shows the differences (in percentage points) of the educational distributions between the CHIP (before and after weighting) and the Census 2010 by age group. The numbers in parentheses are the differences between the weighted CHIP and the Census.

APTable 1.8 Estimated Coefficients of the Impact of Affected Years in the CR on Education Controlling Province Fixed Effects

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.015** (0.006)	-0.011 (0.008)	-0.016* (0.008)	-0.011 (0.011)
Ftype2*YA	-0.002 (0.004)	-0.001 (0.008)	-0.002 (0.005)	-0.004 (0.007)
Ftype3*YA	-0.003 (0.002)	-0.0055** (0.0028)	-0.002 (0.002)	-0.003 (0.003)
YA	-0.001 (0.006)	-0.008 (0.006)	0.003 (0.004)	-0.001 (0.003)
Ftype1	0.370*** (0.018)	0.390*** (0.014)	0.297*** (0.019)	0.294*** (0.024)
Ftype2	0.235*** (0.016)	0.250*** (0.018)	0.126*** (0.011)	0.128*** (0.011)
Ftype3	0.109*** (0.011)	0.115*** (0.014)	0.039*** (0.007)	0.039*** (0.007)
Measure	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
R ²	0.3217	0.3213	0.2615	0.2610
N	32,981	32,981	32,981	32,981

Note: This table reports the regression results for obtaining high school and college degrees based on equation (1.1). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Ftype1 denotes a person's father with high school degree; Ftype2 denotes a person's father has middle school diploma; Ftype3 denotes a person's father has primary school diploma. YA is the measure of years of schooling affected by the CR, using either Meng's measure or Zhou's measure. Control variables include cohort-specific gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects and current residential province fixed effects; The standard errors were clustered at the child's residential province level.

APTable 1.9 Robustness Check: Estimated Coefficients of Equation (1.1), Dropping Elder Sample

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.019** (0.007)	-0.018** (0.008)	-0.018* (0.009)	-0.015 (0.011)
Ftype2*YA	-0.001 (0.004)	-0.002 (0.008)	0.003 (0.005)	-0.004 (0.007)
Ftype3*YA	-0.003 (0.002)	-0.007* (0.003)	-0.001 (0.003)	-0.004 (0.003)
YA	-0.016** (0.005)	-0.014*** (0.003)	-0.00005 (0.003)	-0.002 (0.002)
Ftype1	0.392*** (0.014)	0.390*** (0.014)	0.312*** (0.024)	0.310*** (0.024)
Ftype2	0.247*** (0.019)	0.248*** (0.019)	0.137*** (0.018)	0.141*** (0.018)
Ftype3	0.113*** (0.015)	0.115*** (0.015)	0.041*** (0.008)	0.042*** (0.007)
Measure	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R ²	0.3009	0.3013	0.2485	0.2482
N	30,877	30,877	30,877	30,877

Note: This table lists the regression results for obtaining high school and college degree based on equation (1.1), restricting the sample to people younger than 67 years old. ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. Controls include cohort-variant gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects; the standard errors are clustered at child's residential province level.

APTable 1.10 Estimated Marginal Effects of Equation (1.1), Using Logit and Probit Models

	High School (Logit)		College Degree (Logit)		High School (Probit)		College Degree (Probit)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ftype1*YA	-0.026*** (0.005)	-0.021*** (0.008)	-0.003 (0.004)	-0.001 (0.005)	-0.027*** (0.006)	-0.023*** (0.008)	-0.007* (0.004)	-0.005 (0.005)
Ftype2*YA	-0.001 (0.004)	0.002 (0.006)	0.008** (0.003)	0.006 (0.004)	-0.001 (0.004)	0.001 (0.006)	0.005 (0.003)	0.004 (0.003)
Ftype3*YA	0.001 (0.002)	0.002 (0.003)	0.002 (0.002)	-0.001 (0.003)	0.001 (0.002)	0.001 (0.003)	0.0004 (0.002)	-0.002 (0.002)
YA	0.002 (0.003)	-0.007** (0.003)	-0.003 (0.003)	-0.003 (0.003)	0.002 (0.003)	-0.006** (0.002)	-0.002 (0.002)	-0.001 (0.003)
Ftype1	0.472*** (0.023)	0.464*** (0.023)	0.246*** (0.021)	0.246*** (0.020)	0.485*** (0.022)	0.478*** (0.022)	0.252*** (0.023)	0.251*** (0.022)
Ftype2	0.280*** (0.018)	0.278*** (0.018)	0.155*** (0.015)	0.158*** (0.015)	0.285*** (0.022)	0.284*** (0.019)	0.149*** (0.016)	0.151*** (0.015)
Ftype3	0.141*** (0.009)	0.139*** (0.009)	0.085*** (0.006)	0.087*** (0.006)	0.138*** (0.010)	0.137*** (0.010)	0.077*** (0.006)	0.077*** (0.006)
Measure	Meng	Zhou	Meng	Zhou	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	32,981	32,981	32,981	32,981	32,981	32,981	32,981	32,981
R ²	0.1410	0.1409	0.1904	0.1903	0.1414	0.1413	0.1847	0.1907

Note: This table lists the regression results on obtaining high school and college degree based on equation (1.1), using Logit (Columns (1)-(4)) and Probit (Columns (5)-(8)) models. ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. Controls include cohort-variant gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects; the standard errors are clustered at child's residential province level.

APTable 1.11 Estimated Results of Equation (1.1), Using Schooling Missing Pattern and Years of Father's Schooling

	High School Degree (1)	College Degree (2)
Fschooling*Missing HS	-0.008** (0.004)	-0.011*** (0.003)
Fschooling*Missing HS+MS	-0.012*** (0.004)	-0.013*** (0.003)
Fschooling*Missing HS+MS+PS	-0.013*** (0.004)	-0.013*** (0.004)
Fschooling*Missing MS+PS	-0.007** (0.003)	-0.007*** (0.003)
Fschooling*Missing PS	-0.003 (0.003)	-0.004 (0.003)
Cohort-variant Demog. Var.	Yes	Yes
Cohort FE	Yes	Yes
R ²	0.2980	0.3213
N	32,981	32,981

Note: This table lists the regression results on high school and college degree attainment, using schooling missing pattern dummies as the CR interruption measures, and using years of schooling as the measure of father's education. ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. Schooling missing patterns include missing high school education only, missing high school and middle school, missing high school middle school and primary school, missing middle school and primary school, and missing primary school education only. Controls include cohort-variant gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and the child's birth cohort fixed effects; the standard errors are clustered at child's residential province level.

APTable 1.12 Estimated Results of Equation (1.1), Using Sibling Sample

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.022*** (0.005)	-0.030*** (0.006)	-0.021*** (0.004)	-0.024*** (0.005)
Ftype2*YA	-0.002 (0.004)	-0.007 (0.004)	-0.001 (0.002)	-0.005* (0.003)
Ftype3*YA	-0.001 (0.002)	-0.003* (0.002)	-0.001 (0.001)	-0.002** (0.001)
YA	-0.003 (0.005)	-0.004** (0.002)	0.001 (0.001)	0.001 (0.001)
Ftype1	0.328*** (0.013)	0.329*** (0.013)	0.239*** (0.011)	0.236*** (0.011)
Ftype2	0.168*** (0.010)	0.171*** (0.009)	0.075*** (0.006)	0.077*** (0.006)
Ftype3	0.069*** (0.006)	0.072*** (0.006)	0.021*** (0.003)	0.021*** (0.003)
Measure	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R ²	0.1950	0.1953	0.1644	0.1643
N	54,113	54,113	54,113	54,113

Note: This table reports the regression results for obtaining high school and college degrees based on equation (1.1) using sibling sample. ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. Ftype1 denotes whether a person has a father with high school degree; Ftype2 denotes whether father has middle school diploma; Ftype3 denotes whether father has primary school diploma. YA is the measure of years of schooling affected by the CR, using either Meng's measure or Zhou's measure. Control variables include cohort-variant child's gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects; The standard errors were clustered at the child's residential province level.

APTable 1.13 Estimated Results of Equation (1.1), Adding Family Fixed Effects (Sibling Sample)

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.014* (0.007)	-0.022** (0.010)	-0.014*** (0.004)	-0.016*** (0.005)
Ftype2*YA	-0.006 (0.006)	-0.007 (0.008)	-0.001 (0.002)	-0.004* (0.002)
Ftype3*YA	-0.0002 (0.003)	-0.001 (0.004)	-0.001 (0.001)	-0.002* (0.001)
YA	-0.0003 (0.004)	-0.004 (0.003)	0.002 (0.001)	0.002* (0.001)
Measure	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R ²	0.0309	0.0312	0.0245	0.0246
N	36,361	36,361	36,361	36,361

Note: This table lists the regression results for obtaining high school and college degree based on equation (1.1), adding family fixed effect dummies into the model, using sibling data. ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. Controls include child's cohort-variant gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects; the standard errors are clustered at child's household level.

APTable 1.14 Estimated Results of Equation (1.1), Changing School Entry Age

	High School Degree		College Degree	
	Entry Age: 6	Entry Age: 8	Entry Age: 6	Entry Age: 8
Ftype1*YA	-0.021*** (0.007)	-0.019** (0.007)	-0.022** (0.008)	-0.016* (0.009)
Ftype2*YA	0.001 (0.004)	-0.003 (0.006)	0.006 (0.005)	-0.001 (0.005)
Ftype3*YA	-0.001 (0.002)	-0.006** (0.002)	-0.001 (0.002)	-0.003 (0.003)
YA	-0.003 (0.008)	-0.023*** (0.005)	-0.002 (0.003)	0.0001 (0.002)
Ftype1	0.394*** (0.013)	0.396*** (0.014)	0.314*** (0.022)	0.311*** (0.024)
Ftype2	0.250*** (0.017)	0.254*** (0.017)	0.136*** (0.016)	0.141*** (0.017)
Ftype3	0.112*** (0.013)	0.119*** (0.014)	0.042*** (0.007)	0.045*** (0.008)
Measure	Meng	Meng	Meng	Meng
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R ²	0.3017	0.3024	0.2456	0.2455
N	32,976	32,976	32,976	32,976

Note: This table reports the regression results for obtaining high school and college degrees based on equation (1.1), changing the entry school age in Meng's measure. ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. Ftype1 denotes whether a person has a father with high school degree; Ftype2 denotes whether father has middle school diploma; Ftype3 denotes whether father has primary school diploma. YA is the measure of years of schooling affected by the CR, using either Meng's measure or Zhou's measure. Control variables include child's cohort-variant gender, age and age squared, father's age and father's age squared, and Hukou's type at birth, and child's birth cohort fixed effects; The standard errors were clustered at the child's residential province level.

APTable 1.15 Estimated Coefficients Using All Sample, Urban Sample, and Sample of Urban Residents Who Have Children in the CHIP

	Full Sample (1)	Urban Sample (2)	Urban with Child (3)
Panel A: High School Degree			
FS*CR	-0.009*** (0.002)	-0.003 (0.002)	-0.007* (0.004)
CR	0.142*** (0.021)	0.126*** (0.037)	0.517 (0.319)
FS	0.040*** (0.002)	0.036*** (0.002)	0.039*** (0.004)
N	32,976	10,787	6,182
	0.1587	0.1838	0.1365
Panel B: College Degree			
FS*CR	-0.012*** (0.002)	-0.010*** (0.002)	-0.008*** (0.003)
CR	0.013 (0.016)	-0.019 (0.034)	0.103 (0.193)
FS	0.027*** (0.002)	0.033*** (0.002)	0.025*** (0.003)
N	32,976	10,787	6,182
R ²	0.1456	0.1814	0.0910

Note: This table reports the regression results of the effect of father's schooling on the high school and college degree attainment for the full sample, the urban sample, and the urban sample who had children in the CHIP. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. FS denotes years of father's schooling. CR is a dummy for whether the person received his/her prime education in the CR. The regression also controls for the age, age squared, gender, father's age and father's age squared, and Hukou's type at birth. Standard errors were clustered at child's residential province level.

APTable 1. 16 Decomposition of the Effect of Father's Schooling on College Degree Attainment Using Gelbach (2016)

	Specification		Gelbach (2016) Decomposition		
	Base	Full	Explained FS*YA	Explained YA	Explained FS
FS*YA	-0.0010** (0.0004)	-0.0002 (0.0004)	-0.0008*** (0.0001)	--	--
YA	0.0048*** (0.0009)	0.0009 (0.0008)	--	0.0038*** (0.0005)	--
FS	0.0172*** (0.0017)	0.0070*** (0.0011)	--	--	0.0102*** (0.0008)
HS	No	Yes	-0.00016* (0.00008)	0.00086* (0.00045)	0.00776*** (0.00072)
CEE	No	Yes	-0.00009 (0.00007)	-0.00005 (0.00010)	0.00075*** (0.0006)
CEE score	No	Yes	-0.00058*** (0.00013)	0.00302*** (0.00032)	0.00174*** (0.00053)
Demographic	Yes	Yes			
Cohort FE	Yes	Yes			
Province FE	Yes	Yes			
R ²	0.1648	0.5229			
N	32,981	32,981			

Note: *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The left panel in the table reports the regression results for the base specification and the full specification using equation (6). The base specification does not include HS (whether the person obtained a high school degree), CEE (whether the person took the CEE), and CEEscore (the CEE score divided by 750, the highest possible score), but it includes other demographic variables (age, age squared, gender, father's age and father's age squared, Hukou status at birth), cohort fixed effects and residential province fixed effects of children. The full model specification includes step variables (HS, CEE, CEEscore) and other relevant controls for the CEE scores (year when took CEE, province where took CEE, CEE exam types and the missing value dummies for these variables). The standard errors were clustered at the child's residential province level. The right panel in the table used the decomposition method from Gelbach (2016) to contribute the effect of HS, CEE, and CEEscore to FS*YA, YA, and FS.

APTable 1.17 Occupation as the Measure of Father's Status: Estimated Coefficients of the Effect of Affected Years in the CR on High School and College Degree Attainments

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.014*** (0.007)	-0.026*** (0.008)	-0.024*** (0.005)	-0.035*** (0.008)
Ftype2*YA	-0.014*** (0.006)	-0.018*** (0.007)	-0.019*** (0.005)	-0.020*** (0.006)
Ftype3*YA	-0.010*** (0.003)	-0.015** (0.004)	-0.003 (0.002)	-0.005* (0.003)
YA	0.001 (0.002)	-0.008*** (0.003)	0.004*** (0.001)	0.003* (0.002)
Ftype1	0.479*** (0.014)	0.488*** (0.013)	0.350*** (0.016)	0.348*** (0.016)
Ftype2	0.416*** (0.014)	0.426*** (0.013)	0.350*** (0.016)	0.255*** (0.014)
Ftype3	0.121*** (0.012)	0.128*** (0.011)	0.045*** (0.009)	0.052*** (0.008)
Measure	Meng	Zhou	Meng	Zhou
Cohort-variant Demog. Var.	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
R ²	0.1902	0.1775	0.1673	0.1468
N	31,356	31,356	31,356	31,356

Note: This table reports the regression results for obtaining high school and college degrees based on equation (1.1), using father's occupation as their types. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Ftype1 denotes father's occupation as principal, manager, technician or professional; Ftype2 includes clerk, commercial and service personnel; Ftype3 are farmers or manufacturing related personnel. Controls include child's cohort-variant gender, age and age squared, father's age and father's age squared, Hukou type at birth, and child's birth cohort fixed effects; the standard errors are clustered at the child's residential province level.

APTable 1.18 Occupation as Measure of Father's Status: Estimated Coefficients of the Impact of the CR on Multi-Generational Mobility for High School and College Degree

	High School			College		
	(1)	(2)	(3)	(4)	(5)	(6)
GFtype1*YA	-0.004 (0.011)	-	-0.004 (0.012)	0.009 (0.018)	-	0.007 (0.021)
GFtype2*YA	-0.020** (0.007)	-	-0.020** (0.007)	-0.013** (0.004)	-	-0.015** (0.006)
GFtype3*YA	0.011 (0.044)	-	0.005 (0.011)	0.007 (0.012)	-	0.002 (0.011)
Ftype1*YA	-	0.004 (0.009)	0.017 (0.009)	-	0.006 (0.011)	0.016 (0.011)
Ftype2*YA	-	0.022 (0.017)	0.0010 (0.008)	-	0.020* (0.010)	0.021* (0.010)
Ftype3*YA	-	0.022 (0.017)	0.018 (0.018)	-	0.013 (0.011)	0.007 (0.012)
YA	-0.003 (0.010)	-0.028 (0.050)	-0.003 (0.010)	-0.005 (0.009)	-0.003 (0.008)	-0.004 (0.008)
GFtype1	0.154*** (0.043)	-	0.110** (0.047)	0.186*** (0.075)	-	0.147* (0.083)
GFtype2	0.184*** (0.040)	-	0.088** (0.036)	0.226*** (0.049)	-	0.214*** (0.053)
GFtype3	-0.009 (0.044)	-	0.001 (0.042)	0.032 (0.062)	-	0.041 (0.053)
Ftype1	-	0.152*** (0.043)	0.179*** (0.041)	-	0.190*** (0.038)	0.163*** (0.037)
Ftype2	-	0.100** (0.035)	0.088** (0.036)	-	0.059 (0.035)	0.050 (0.035)
Ftype3	-	-0.121 (0.069)	-0.105 (0.074)	-	-0.072 (0.050)	-0.048 (0.052)
N	4,684	4,684	4,684	4,684	4,684	4,684
R ²	0.0959	0.1202	0.1346	0.1382	0.1468	0.1708

Note: This table reports the regression results for obtaining high school and college degrees based on equations (1.3)-(1.5), using father's occupation as their types. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Controls include child's cohort-variant gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects; the standard errors are clustered at the child's residential province level.

APTable 1.19 Estimated Coefficients of the Effect of Mother's Schooling and Father's Schooling on High School and College Attainments: Based on Equations (1.7), (1.8) and (1.9)

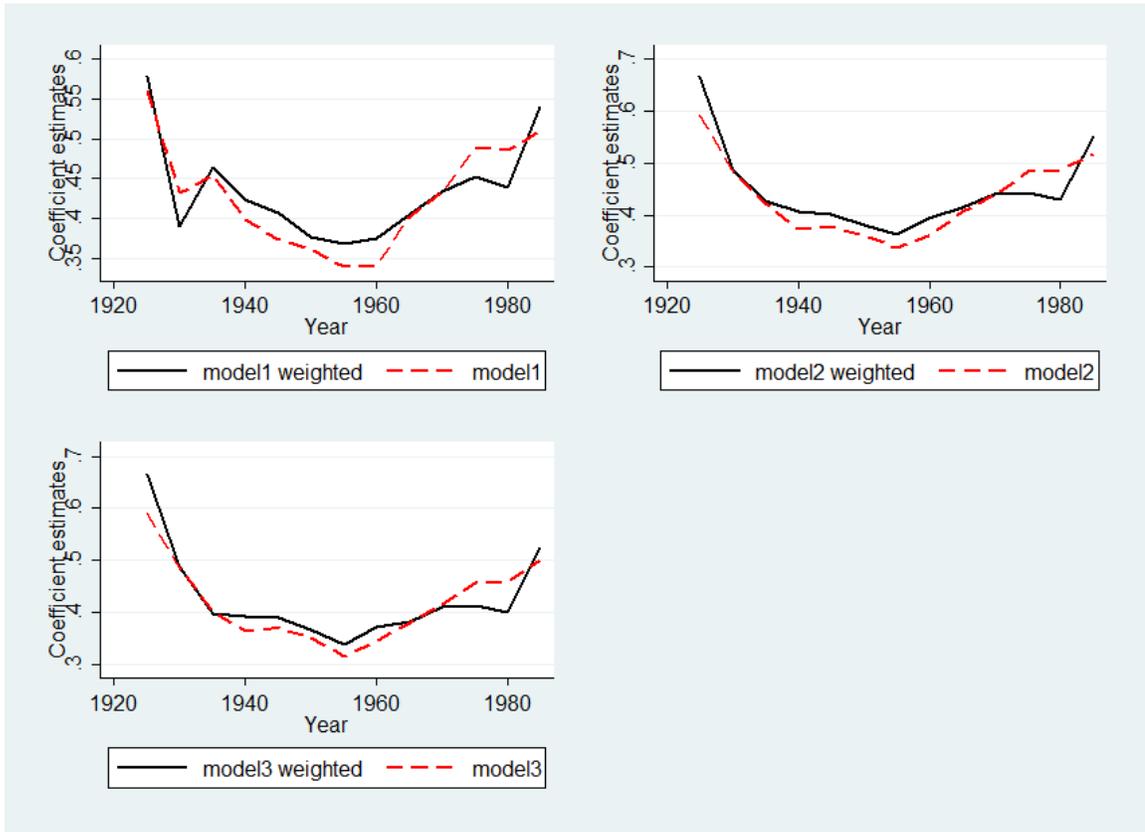
	High School			College		
	(1)	(2)	(3)	(4)	(5)	(6)
MS*YA	-0.0001 (0.0005)	-	-0.0001 (0.0001)	-0.0011** (0.0005)	-	-0.0004 (0.0005)
FS*YA	-	- 0.0013*** (0.0004)	-0.0009** (0.0004)	-	- 0.0020*** (0.0003)	- 0.0015*** (0.003)
YA	0.012*** (0.002)	0.016*** (0.002)	-0.003 (0.002)	0.005*** (0.001)	0.010*** (0.001)	0.005*** (0.001)
MS	0.044*** (0.001)	-	0.026*** (0.001)	0.031*** (0.001)	-	0.020*** (0.001)
FS	-	0.042*** (0.001)	0.029*** (0.001)	-	0.030*** (0.001)	0.018*** (0.001)
N	31,205	31,337	31,187	31,205	31,337	31,187
R ²	0.1748	0.1828	0.2065	0.1660	0.1658	0.1879

Note: This table reports the regression results for obtaining high school and college degrees based on equations (1.7)-(1.9). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. MS denotes years of mother's schooling; FS denotes years of father's schooling; YA denotes the years a person was affected by the CR. Controls in columns (1) and (4) include child's cohort-variant gender, age and age squared, mother's age and mother's age squared, Hukou's type at birth, and child's birth cohort fixed effects; controls in columns (2) and (5) include cohort-variant gender, age and age squared, father's age and father's age squared, Hukou's type at birth, and child's birth cohort fixed effects; controls in columns (3) and (6) include all of the controls mentioned above. The standard errors are clustered at the child's residential province level.

APTable 1.20 Sample Descriptive Statistics by the Answer of "Whether Took CEE" in the CHIP

	Pre-CR	CR	Post-CR
(1) Whether Took CEE			
	N (%)	N (%)	N (%)
No	616(29%)	4,098(31%)	5,247(30%)
Yes	56(3%)	778(6%)	2,499(14%)
Missing	1,427(68%)	8,508(63%)	9,747(55%)
Total	2,099(100%)	13,384(100%)	17,493(100%)
(2) Whether Took CEE among People with Years of Schooling>12			
	N (%)	N (%)	N (%)
No	231(75%)	1,398(65%)	2,564(48%)
Yes	56(18%)	584(27%)	2,499(47%)
Missing	22(7%)	178(8%)	297(6%)
Total	308(100%)	2,160(100%)	5,323(100%)
(3) Descriptive Statistics by Whether Took CEE among People with Years of			
	Mean (Std)	Mean (Std)	Mean (Std)
(i) Age			
CEE taker	73.38(4.43)	54.70(3.82)	28.25(6.87)
No taker	73.41(4.23)	57.21(4.21)	40.13(6.36)
Missing	73.25(5.03)	56.42(3.96)	40.78(6.35)
(ii) Schooling			
CEE taker	15.45(1.80)	13.49(1.83)	14.78(1.93)
No taker	12.90(1.70)	12.53(1.44)	12.80(1.70)
Missing	12.60(1.67)	12.29(1.17)	12.40(1.35)
(iii) Father's Schooling			
CEE taker	5.45(4.87)	6.00(4.50)	8.59(4.16)
No taker	4.00(4.12)	4.70(4.30)	7.39(4.00)
Missing	3.76(4.16)	4.03(3.89)	6.82(3.82)
(iv) Mother's Schooling			
CEE taker	2.92(4.05)	4.05(4.38)	6.96(4.45)
No taker	1.59(3.10)	2.68(3.79)	5.56(4.11)
Missing	1.42(2.71)	2.37(3.37)	5.55(3.83)

APFigure 1.1 Estimated Coefficients of the Intergenerational Coefficients of APTable 1.2 (Rank of Schooling)



Note: These graphs were plotted based on the estimated results of the three model specifications in APTable 1.2. Model 1 is the specification with no controls, model 2 has basic controls, and model 3 has extensive controls. The solid lines are estimated intergenerational coefficients using the weights described in Appendix A1, and the dash lines are estimated coefficients without weighting.

APFigure 1.2 Estimated Coefficients of the Intergenerational Correlations of APTable 1.3 (Years of Schooling)



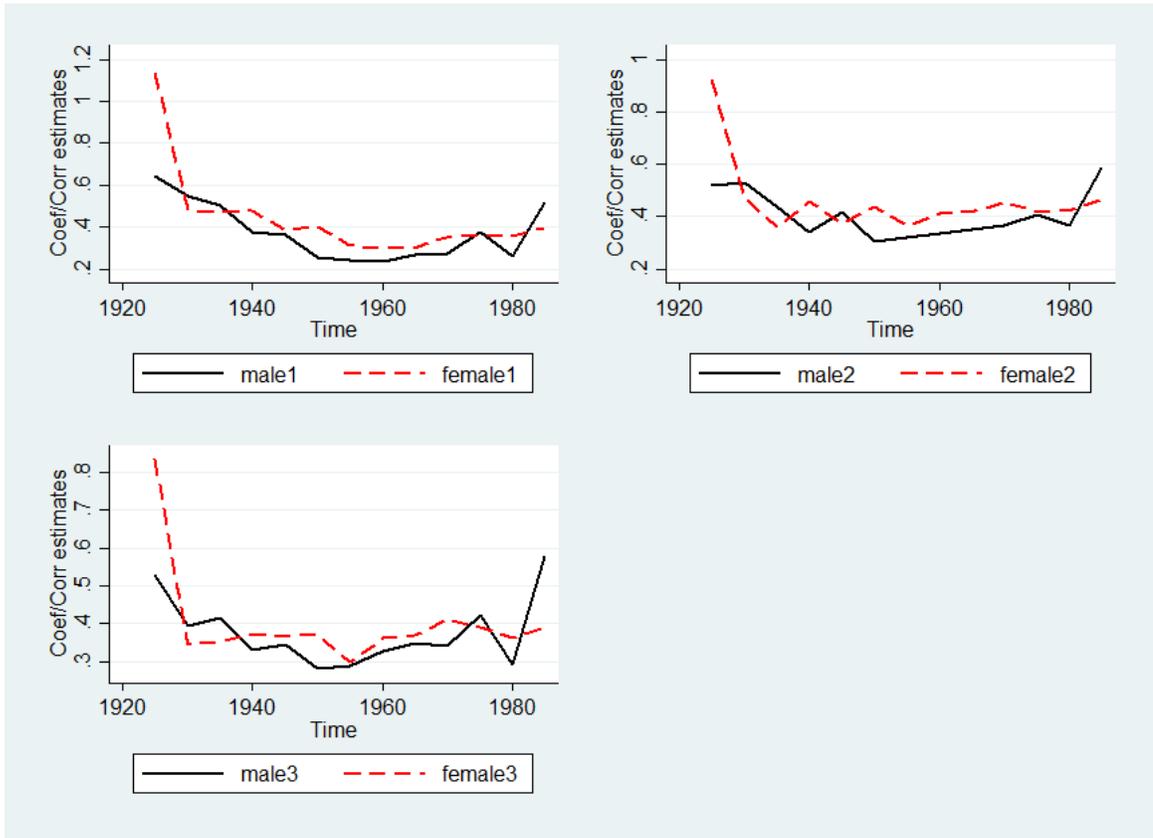
Note: These graphs were plotted based on the estimated results of the three model specifications in APTable 1.3. Model 1 is the specification with no controls, model 2 has basic controls, and model 3 has extensive controls. The solid lines are estimated intergenerational coefficients using the weights described in Appendix A1, and the dash lines are estimated coefficients without weighting.

APFigure 1.3 Estimated Coefficients of the Intergenerational Schooling Coefficients and Correlations for Urban and Rural Residents



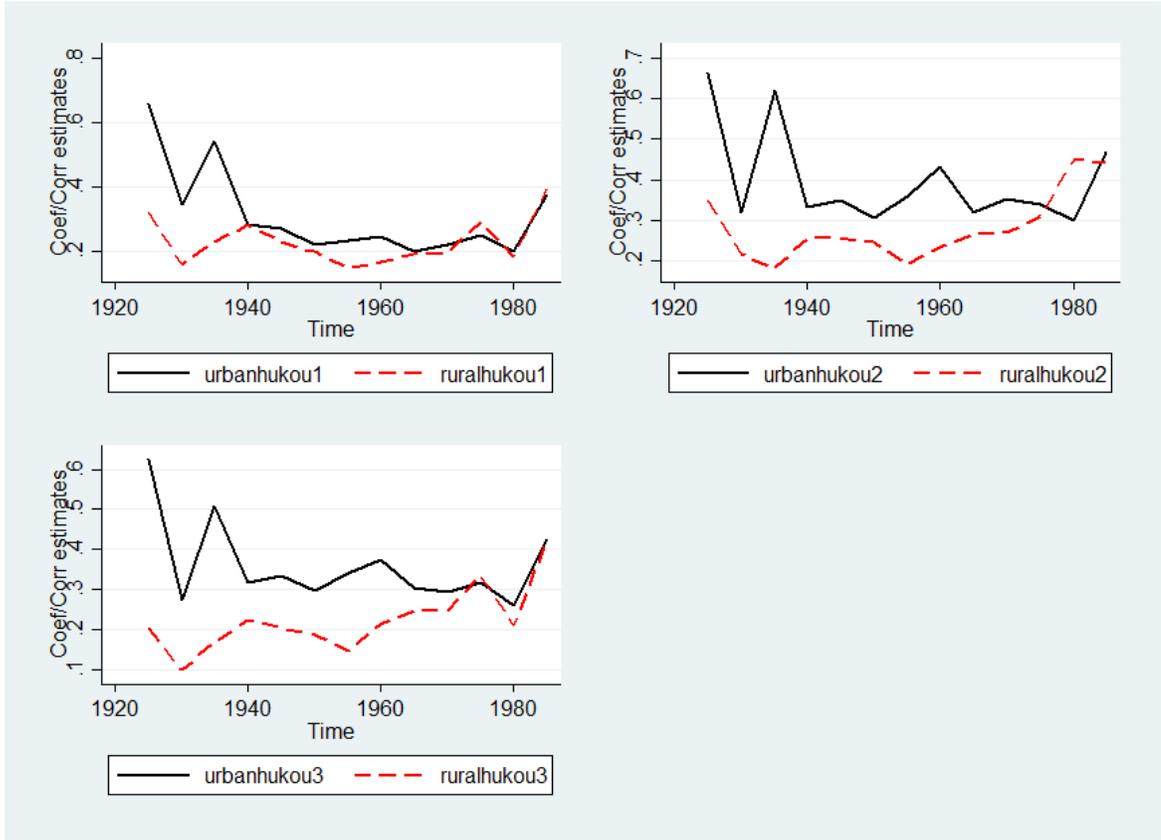
Note: These graphs were plotted based on the results using the model in Chen et al. (2015) and divided the sample into rural and urban samples. The top-left subgraph plots the estimates of the intergenerational coefficients using years of schooling, the top-right subgraph plots the estimates of the intergenerational coefficients using rank of schooling, and the bottom-left subgraph plots intergenerational correlations using years of schooling. The solid lines are the estimates for urban sample; the dotted lines are the estimates for rural sample.

APFigure 1.4 Estimates of the Intergenerational Schooling Coefficients and Correlations for Male and Female



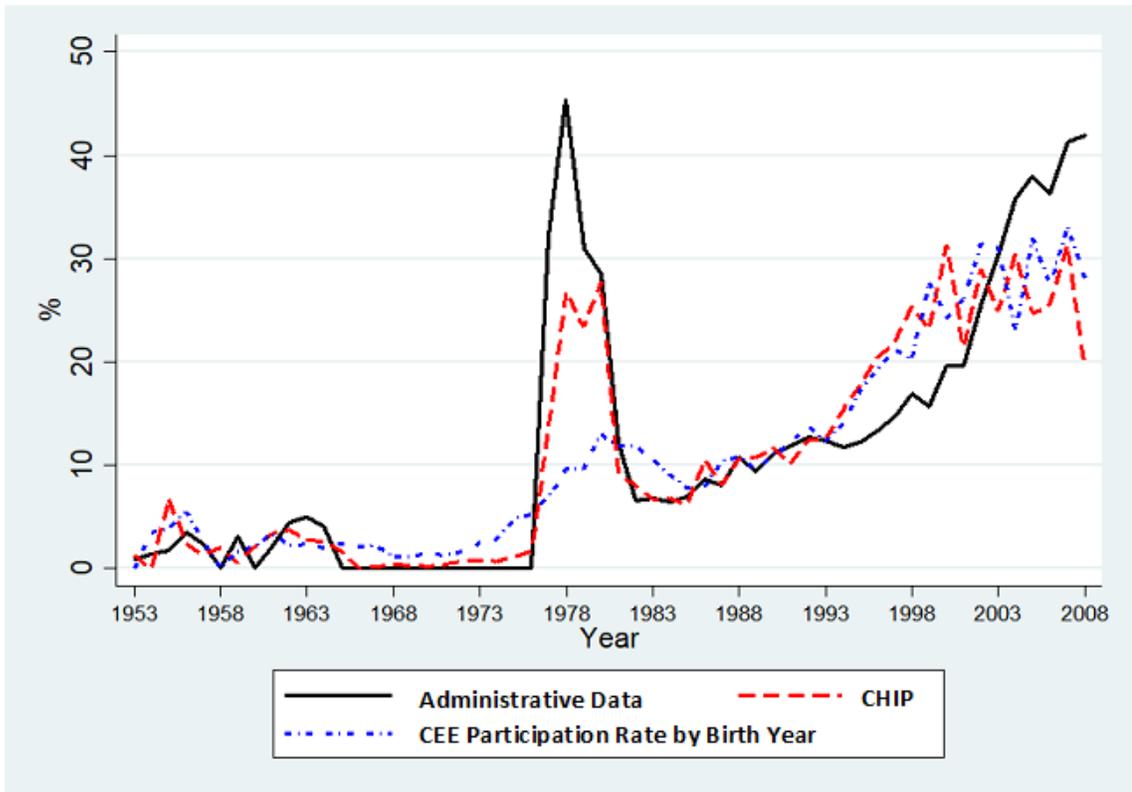
Note: These graphs were plotted based on the results using the model in Chen et al. (2015) and divided the sample into male and female samples. The top-left subgraph plots the estimates of the intergenerational coefficients using years of schooling, the top-right subgraph plots the estimates of the intergenerational coefficients using rank of schooling, and the bottom-left subgraph plots intergenerational correlations using years of schooling. The solid lines are the estimates for male sample; the dotted lines are the estimates for female sample.

APFigure 1.5 Estimates of the Intergenerational Schooling Coefficients and Correlations for Urban-born and Rural-born Samples



Note: These graphs were plotted based on the results using the model in Chen et al. (2015) and divided the sample into agricultural (rural) Hukou at birth and non-agricultural (urban) at birth samples. The top-left subgraph plots the estimates of the intergenerational coefficients using years of schooling, the top-right subgraph plots the estimates of the intergenerational coefficients using rank of schooling, and the bottom-left subgraph plots intergenerational correlations using years of schooling. The solid lines are the estimates for urban-born sample (proxied by people who had an Urban Hukou at birth), the dotted lines are the estimates for rural-born sample (proxied by people who had an Agricultural Hukou at birth).

APFigure 1.6 Number of People Who Reported Taking the CEE as Proportions of 19-year-old Population and the Proportion of 19-year-old Population Registered for the CEE by Year



Note: This graph plots the fraction of CEE takers as a proportion of 19-year-olds, comparing the CHIP and administrative data by year. The solid line shows the fractions of CEE takers using Census 2000 and an administrative file of CEE takers by CEE-taking year, the short-dashed line shows the fractions of CEE takers reported by CHIP respondents by CEE-taking year, and the dotted line shows the proportion of 19-year-olds who took the CEE by their birth year (the horizontal line denotes the year when they were 19 years old) using the CHIP.

1.10 Reference

Andreas, Joel. *Rise of the Red Engineers: The Cultural Revolution and the Origins of China's New Class*. Stanford University Press, 2009.

Becker, Gary S., and Nigel Tomes. "Child Endowments and the Quantity and Quality of Children." *Journal of political Economy* 84.4, Part 2 (1976): S143-S162.

Becker, Gary S., and Nigel Tomes. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4.3, Part 2 (1986): S1-S39.

Behrman, Jere, and Paul Taubman. "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model." *The Review of Economics and Statistics* (1985): 144-151.

Behrman, Jere R., and Mark R. Rosenzweig. "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" *American Economic Review* 92.1 (2002): 323-334.

Blanden, Jo, and Stephen Machin. "Educational Inequality and the Expansion of UK Higher Education." *Scottish Journal of Political Economy* 51.2 (2004): 230-249.

Björklund, Anders, Markus Jantti, and Gary Solon. "Influences of Nature and Nurture on Earnings Variation." *Unequal Chances: Family Background and Economic Success*. Princeton University Press, 2005. 145-164.

Bratberg, Espen, Øivind Anti Nilsen, and Kjell Vaage. "Intergenerational Earnings Mobility in Norway: Levels and Trends." *Scandinavian Journal of Economics* 107.3 (2005): 419-435.

Chen, Yuyu, Suresh Naidu, Tinghua Yu, and Noam Yuchtman. "Intergenerational Mobility and Institutional Change in 20th Century China." *Explorations in Economic History* 58 (2015): 44-73.

[dataset] Chinese Household Income Project, 2013 (2014). China Institute for Income Distribution. Applied and Retrieved from China Institute for Income Distribution website: <http://www.ciidbnu.org/chip/chips.asp?year=2013>.

Chetty, Raj, et al. "Where is the land of opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129.4 (2014): 1553-1623.

Clark, Gregory. *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press, 2015.

Corak, Miles. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27.3 (2013): 79-102.

Dahl, Gordon B., and Lance Lochner. "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review* 102.5 (2012): 1927-56.

Deng, Zhong, and Donald J. Treiman. "The Impact of the Cultural Revolution on Trends in Educational Attainment in the People's Republic of China." *American Journal of Sociology* 103.2 (1997): 391-428.

Du, Fang Cai Yang. "Destructive Effects of Cultural Revolution on Physical and Human Capital." *China Economic Quarterly* 3 (2003): 2.

Giles, John, Albert Park, and Meiyan Wang. "The Great Proletarian Cultural Revolution, Disruptions to Education, and the Returns to Schooling in Urban China." *Economic Development and Cultural Change* 68.1 (2019): 131-164.

Han, Dongping. "Impact of the Cultural Revolution on Rural Education and Economic Development: The Case of Jimo County." *Modern China* 27.1 (2001): 59-90.

Heckman, James J., and V. Joseph Hotz. "An Investigation of the Labor Market Earnings of Panamanian Males: Evaluating the Sources of Inequality." *Journal of Human Resources* (1986): 507-542.

Hertz, Tom, et al. "The Inheritance of Educational Inequality: International Comparisons and Fifty-year Trends." *The BE Journal of Economic Analysis & Policy* 7.2 (2008).

Jantti, Markus, et al. (2006) "American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States." IZA Discussion Paper No. 1938, Available at: <https://www.iza.org/publications/dp/1938/american-exceptionalism-in-a-new-light-a-comparison-of-intergenerational-earnings-mobility-in-the-nordic-countries-the-united-kingdom-and-the-united-states>.

Lee, Chul-In, and Gary Solon. "Trends in Intergenerational Income Mobility." *The Review of Economics and Statistics* 91.4 (2009): 766-772.

Lucas, Robert EB, and Sari Pekkala Kerr. "Intergenerational Income Immobility in Finland: Contrasting Roles for Parental Earnings and Family Income." *Journal of Population Economics* 26.3 (2013): 1057-1094.

Ma, Jiantang, et al. (2014) *China Statistical Yearbook 2014*. China Statistics Press.

Mazumder, Bhashkar. "Sibling Similarities and Economic Inequality in the US." *Journal of Population Economics* 21.3 (2008): 685-701.

Meng, Xin, and Robert G. Gregory. "The Impact of Interrupted Education on Subsequent Educational Attainment: A Cost of the Chinese Cultural Revolution." *Economic Development and Cultural Change* 50.4 (2002): 935-959.

Meng, Xin, and Robert Gregory. (2007) *Exploring the Impact of Interrupted Education on Earnings: The Educational Cost of the Chinese Cultural Revolution*. IZA Discussion Paper No. 2548. Available at: <https://www.iza.org/publications/dp/2548/exploring-the-impact-of-interrupted-education-on-earnings-the-educational-cost-of-the-chinese-cultural-revolution#:~:text=2548%3A%20Exploring%20the%20Impact%20of,of%20the%20Chinese%20Cultural%20Revolution&text=We%20estimate%20that%20those%20who,50%20percent%20of%20potential%20earnings.>

[dataset] Office of Population Census at the State Council of P.R.C. and Department of Population and Employment Statistics National Bureau of Statistics of P.R.C. (2000, 2010). Population Census of the P.R.C. [Electronic Version]. Retrieved from China Statistics Bureau Database. <http://www.stats.gov.cn/tjsj/pcsj/rkpc/5rp/index.htm>.

Ridge, John M. "Three Generations." *Mobility in Britain Reconsidered* (1973): 47-71.

Roland, Gerard, and David Y. Yang. (2017) *China's Lost Generation: Changes in Beliefs and Their Intergenerational Transmission*. No. w23441. National Bureau of Economic Research. Available at: <https://www.nber.org/papers/w23441>.

Schultz, T. Paul. "Human Capital, Family Planning, and Their Effects on Population Growth." *The American Economic Review* 84.2 (1994): 255-260.

Solon, Gary. "A Model of Intergenerational Mobility Variation over Time and Place." *Generational Income Mobility in North America and Europe 2* (2004): 38-47.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. "What Are We Weighting For?." *Journal of Human Resources* 50.2 (2015): 301-316.

Solon, Gary. "What Do We Know So Far about Multigenerational Mobility?" *The Economic Journal* 128.612 (2018): F340-F352.

Stuhler, Jan. (2012) *Mobility across Multiple Generations: The Iterated Regression Fallacy*. IZA Discussion Paper No. 7072. Available at: <https://www.iza.org/publications/dp/7072/mobility-across-multiple-generations-the-iterated-regression-fallacy>.

Thomas, Duncan, Robert F. Schoeni, and John Strauss. (1996) *Parental Investments in Schooling: The Roles of Gender and Resources in Urban Brazil*. Labor and Population Program, No. 96-02. Available at: <https://www.rand.org/pubs/drafts/DRU1303.html>.

Warren, John Robert, and Robert M. Hauser. "Social Stratification across Three Generations: New evidence from the Wisconsin Longitudinal Study." *American Sociological Review* (1997): 561-572.

Xie, Yu, and Chunni Zhang. "The long-term impact of the Communist Revolution on social stratification in contemporary China." *Proceedings of the National Academy of Sciences* 116.39 (2019): 19392-19397.

Zeng, Zhen, and Yu Xie. "The Effects of Grandparents on Children's Schooling: Evidence from Rural China." *Demography* 51.2 (2014): 599-617.

Zhang, Junsen, Pak-Wai Liu, and Linda Yung. "The Cultural Revolution and Returns to Schooling in China: Estimates Based on Twins." *Journal of Development Economics* 84.2 (2007): 631-639.

Zhou, Dong. "The Long Term Impacts of the Cultural Revolution on Economic Performance of Urban Residents in China." Unpublished, University of California, Riverside September (2014).

Chapter 2

Impact of the College Expansion on Skilled Labor Market in China

2.1 Introduction

The Chinese higher education system had been suspended during the Cultural Revolution (1966-1976) and started to recover and gradually expand beginning 1978. However, the growth was slow, and in 1999, fewer than 10% of those age 18-22 were enrolled in colleges and few people in the employed labor force had college degrees (3.5% of the employed workers had college degrees in 1998). Starting from 1999, China began an unprecedented expansion of its higher education system. According to official descriptions, the Chinese government adopted suggestions of Chinese economist Min Tang to double the college admission in China in 4 years and maintain a high annual rate of increase in college admission after that point (Tang & Zuo, 1999).

According to basic economic theories, a large increase in college admission will bring more college graduates to the labor market, which creates a positive supply shock and affects labor market outcomes. Many papers have studied the impact of the increase in number of college graduates in the early 1970s in the US affected the labor market outcomes. Freeman (1975) found the increase in supply of college graduate workers relative to demand in the U.S during 1970s reduced the relative wage of college graduates, reduced the rate of return to investing higher education, and reduced employment

opportunity for college graduates. Freeman (1977) further showed that the negative impact was more serious for young workers than older workers. Goldin et al. (2007) also attributed the decline in relative wage between college and high school graduates in the U.S in 1970s to the increase of relative supply. More empirical analysis on the impact of the increase in college graduates in the US can also be found in Welch (1979) and Katz and Murphy (1992). Aamodt and Arnesen (1995) found empirical evidence in Norway that the college expansion starting in 1989 in Norway increased the likelihood of unemployment for college graduates six months after their graduation.

In the long run, it has been argued that a positive supply shock of skilled labor can also stimulate companies to develop skill-complementary technology and thus increase the productivity as well as the premium for skill labor (Acemoglu, 1998, 2007). Carneiro et al. (2018) find the college expansion in Norway in 1970s increased the productivity of skilled labor and investments in R&D.

Some studies have discussed the impact of the college expansion in China on labor market outcomes. Several studies have focused on the short-run market responses to the college expansion. Li et al. (2014) used 2000 and 2005 1% population survey data and found that the college expansion increased the probability of college attendance among high school graduates and increased the unemployment rate among young college graduates. He et al. (2015) conducted a descriptive study based on administrative data from the Chinese higher education system and labor market. He argued that the quality of a college education decreased due to the college expansion, which may have caused college graduates to find lower-paid jobs or become unemployed. Knight et al. (2017) summarized the relative wages, probability of unemployment, and likelihood of obtaining “good jobs”

(technical, professional, government or enterprise management) for college graduates using survey data from 2002 to 2008. These papers studied the impact of the college expansion either in descriptive ways (He et al., 2015, Knight et al., 2017), or compared single-year outcomes before and after the college expansion (Li et al., 2014, Knight et al., 2017). They did not develop measures to quantify the college expansion, and, as a result, they are not able to provide precise estimates of the impact of the college expansion on labor market outcomes.

Xing et al. (2018) studied the mid-run impact of the college expansion on the probability of unemployment for college graduates. Their study used census data for 2000, the 1% population survey for 2005, and the census data for 2010. Using a difference-in-difference identification strategy, they found that although the unemployment rate for young people increased between 2000 to 2005, the likelihood of unemployment for young people in 2010 was not statistically different from 2000. Yao (2015) use municipal level data in China to study the long-term impact of college expansion on regional economies. His results suggest the increase in college construction had a positive effect on regional average wage and real GDP per capita in the long term.

In this chapter, we study the short run impact of the college expansion on the college premium, unemployment, and the skills used in first jobs in China. Our paper contributes to the previous literature in several ways.

First, we use the number of students admitted to college to quantify the college expansion and study its impacts on labor market outcomes for younger and older workers. We interpret our estimated results as the impact of a positive supply shock on the labor

market. Since the increased rate of college admission is based on pre-established government policies, our measure of the supply shock in skilled workers allows us to separate supply and demand effects when on the college premium.

Second, we study the impact of the college expansion on the college premium within industry and occupation. By controlling industry fixed effects and occupation fixed effects, we identify a decline in the college premium within industry and occupation due to the college expansion.

Third, we study the impact of the college expansion on skills used in a first job. We combine individual occupation data from the CGSS with O*NET codes to calculate the scores of cognitive and manual tasks used in first jobs. This analysis suggests that the decrease in the college premium may also be attributed to the movement of skilled labor between occupation.

This chapter is structured as follows: Section 2.2 introduces the background for the college expansion and considers the role of college graduate workers in the labor market. Section 2.3 describes the data we use. Section 2.4 describes the identification strategy used in this chapter. Section 2.5 reports the estimated impact of college expansion on the college premium, unemployment, and types of tasks used in first jobs. Section 2.6 summarizes the chapter.

2.2 Background

2.2.1 The College Expansion in China

College enrollment in China has increased substantially since the Cultural Revolution, especially after 1999. The average annual increase rate for the number of students admitted

to colleges was around 8.5% from 1978 to 1998. In 1999, the Chinese government announced a decision to accelerate the expansion of higher education in order to increase human capital accumulation, stimulate domestic demand, and improve employment options for young people and workers laid off from State Owned Enterprises. The college admissions increased at an annual rate of more than 10% from 1999 to 2012. This period of extraordinary growth came to an end in 2012, when the Ministry of Education of China (MOE) issued guidelines slowing the growth of higher education enrollment, indicating the need to guarantee quality of instruction (MOE, 2012).

Figure 2.1 shows the number of students admitted to regular and adult 4-year colleges¹⁶ and 2- or 3-year colleges¹⁷ in China from 1992 to 2015. The number of students admitted to regular colleges (sum of the line with triangle markers and the line with round markers) increased by more than 45% in 1999, and the college admission grew at an average rate of 20% per year from 1999 to 2008. In 2008, 6 million students were admitted to regular colleges, about 5.5 times the number in 1998. The level of admission to 2- or 3-year colleges was similar to that for 4-year colleges before the college expansion, but it increased at a lower rate than 4-year college admission in the beginning years of the college expansion.

Although the number of students admitted to regular colleges was about the same as the number of students admitted to adult colleges before the college expansion, the

¹⁶ The 4-year colleges are also called the undergraduate colleges, and students receive bachelor's degrees. Students who graduate from 2 or 3-year colleges receive associate's college degrees.

¹⁷ Regular higher education institutions provide full-time higher education instruction. It includes 4-year universities and colleges, 2 or 3-year colleges, vocational colleges, and independent colleges (see footnote 1). Adult higher education institutions mainly admit part-time adults based on the adult college entrance exam.

admissions in regular colleges increased much faster than the admissions in adult colleges during the college expansion period. In 2008, the number of admissions into adult college was only about one-third of the admissions into regular colleges. The 2- or 3-year adult college admissions increased in the beginning of the college expansion but decreased by 540 thousand from 2002 to 2006, followed by a modest growth. The admissions for 4-year adult colleges were less than 12,000 up through 1998, and although proportional growth was large after that point, the total contribution to enrollment was modest compared to that of regular 4-year colleges.

After 2008, the government slowed down the expansion rate for college admission. The annual growth rate for overall admissions into colleges (both regular and adult) shrank to about 3% between 2008 to 2015.

The growth rates of college admissions differ across provinces during the college expansion period. Figure 2.2 shows the average annual increase rate of college admissions in each province of China from 1998 to 2012¹⁸. Most provinces increase at more than 10% per year from 1998 to 2012. Provinces with the highest density of higher education initially, like Beijing and Shanghai, along with other provinces in northeastern and northwestern areas have the lowest rate of increase in college admissions. Less-developed provinces in general have higher rates of increase than more developed provinces.

The proportion of the population age 18-22 enrolled in colleges (both regular and adult colleges) increased from 9.8% in 1998 to 23.3% in 2008 (Figure 2.3, the circle line). The

¹⁸ The equation to calculate the average annual increase rate of college admissions is: $IR = (\sqrt[14]{AD_{2012}/AD_{1998}} - 1) * 100\%$, where $AD_{p,t}$ is the number of college admissions in province p in year t .

18-22 year-old population (Figure 2.3, the cross line) increased at a rate of 5% in 1998-2009 and decreased at 6% per year in 2009-2015, a result of China's one-child policy. Thus, it suggests that the college admissions increased at a faster rate than the 18-22 population in 1998-2009, and the continued growth in proportion of college enrollment among those aged 18-22 in 2010-2015 reflects both the continuous declines in the 18-22 population and increases in the number of students attending college.

Although college admissions started to surge in 1999, the funding, facilities and other resources for higher education did not increase proportionally in the beginning of the college expansion. Figure 2.4 shows the national public budget levels for regular higher education institution, as well as the student-teacher ratio¹⁹ for 4-year and 2- or 3-year regular colleges in 1996-2015. As can be seen from the graph, the budget (the cross line) increased almost linearly at an annual rate of 12.7% from 1999 to 2008 but grew faster (15.6% per year) from 2008 to 2015. Since the annual growth rate for regular college admissions (20%) was much higher than the growth rate for the public budget, the student-teacher ratio (circle line) increases dramatically during the initial period of college expansion.

For adult higher education institutions (Figure 2.5), public budget and student-teacher ratio both increased in 1996-2015. It is worth noting that the public budget allocated to adult colleges was much smaller than that to the regular colleges, especially after the college expansion. The adult college budget was only about 1/66 of the budget level for regular colleges in 2015, although the adult college admissions was about 1/3 of the

¹⁹ The student-teacher ratio is calculated as number of students in colleges divided by the number of full-time teachers in colleges.

admission in regular colleges. The student teacher ratio is more than ten times that for regular colleges. Thus, in the college expansion, the government allocated more resources to regular colleges than adult colleges, although the admissions into adult colleges also substantially increased.

The number of 2- or 3-year colleges began to increase with the college expansion and had tripled by 2008, as shown in Figure 2.6 (cross line), while the number of 4-year colleges (circle line) increased only slightly in the beginning of the college expansion but then increased by a more than 45% in 2008. The sudden increase of 4-year colleges in 2008 is a result of increase in dependent colleges²⁰ which are private 4-year colleges affiliated with regular public colleges. The 2- or 3-year colleges increased in an annual rate of 16% from 1999 to 2005 but only 2.1% per year after 2005.

Accompanying the college expansion, MOE decentralized its control of many colleges to province governments and allowed more private funded colleges to be established. This led to a large increase in the number of colleges under the provincial governments (triangle line in Figure 2.7) and a decrease in the number of colleges under the central government (circle line in Figure 2.7). Also, the number of private colleges²¹ (cross line in Figure 2.7)

²⁰ The independent colleges in China are 4-year institutions that were created from units of regular colleges and funded jointly by non-government organizations or individuals and regular colleges. The *Measures for the Establishment and Administration of Independent Colleges* was published in 2008 (MOE, 2008), which officially established the legal status of independent colleges.

²¹ Private colleges in China are either 2, 3 or 4-year colleges, which are funded by non-public sources. The average tuition for private colleges is 3-4 times the average tuition for public colleges. Private colleges are under the control of provincial ministry of education. Due to lack of separate statistics for private colleges before 2002 in the Education Yearbook, we plot the number of private colleges in 2002-2015. The number of private colleges before 2002 is included in the number of colleges under provincial governments (triangle line in Figure 2.7).

grew starting in 2001 and surged in 2008 due to the creation of independent colleges, as noted above.

In summary, the unprecedented expansion in Chinese higher education has greatly increased the chance that young people receive a college education. College admission for both 4-year and 2- or 3-year colleges has increased at a very high rate in 1999-2012. Although the number of colleges, number of college teachers and national expenditures on higher education all increased after 1999, they did not increase proportionately with the college admission in the beginning of the college expansion. The major expansion happened in 2- or 3-year colleges, locally controlled colleges and private colleges. The lack of resources for college expansion at the beginning of the period of growth, and the unbalanced expansion toward less elite colleges may have resulted in a decrease to the quality of higher education.

2.2.2 College Admission Planning in China

Since the foundation of the PRC, the Ministry of Education in China has coordinated with national economic planning departments to schedule the number of college admissions for central-government-controlled colleges for each province in each year. In the period 1999-2012, these planning numbers were made according to the five-year or ten-year plan for higher education, taking into account the plans for socioeconomic development of each province in each year.

Specifically, the enrollment plan made by the central departments was distributed to the department of education in each province. Based on the planned number of college

admissions, the department of education in each province made enrollment plans for each college under the control of the provincial government and located within the province.

After discussing the total number of admissions with the provincial department of education, each school reports its plans back to the department of education in each province and publishes their admission plans to the public before the administration of the college entrance exam in each year. After receiving the plan from schools, the department of education in each province reviews them and summarizes the total number of college admissions within the province. Then, they report back the calculated numbers to the central ministry of education. After receiving the confirmation from the central ministry, these admission plans at the provincial level are also published. Therefore, during the college expansion period, the increase in number of college admissions of each province in each year is pre-planned by the Ministry of Education in China and the department of education in each province based on targeted increase rates in college admissions, which differ by province, as well as the economic development plans for each province.

2.2.3 College Graduates in the Chinese Labor Market

The large increase in the number of students admitted to college will increase the supply of skilled labor in the market, and we expect it may change the relative price and employment status for skilled workers.

Figure 2.8 shows the educational distribution for employed workers in 1996 to 2015 according to China Labor Statistical Yearbook. The area from the top to the bottom represents the proportion of workers whose highest education is at least college degree, high school degree, middle school, and primary school degree (6-year or 5-year), and less

than a primary school degree, respectively. As can be seen from the graph, the proportion of college graduate among all workers has increased from 2.8% in 1996 to 17.5% in 2015, while the proportion with a primary education or less declined from 48.3% in 1996 to 20.6% in 2015. This implies an overall skill upgrading among employed workers in China.

Figure 2.9 shows the changes of proportions of college worker proportions from 1998 to 2012 by province. We can see that the proportions of college graduate workers in most provinces increase by 10-15 percentage points in 1998-2012, indicating a large upgrading in labor force in most provinces. The increase of the proportions of college worker in the south-western area, where the least-developed provinces are located, are the smallest. The greatest increase occurred in the most-developed provinces, like Beijing, Shanghai, Jiangsu, and Zhejiang. Comparing Figure 2.9 with the observed pattern for the average annual increases of college admissions in Figure 2.2 where less-developed provinces experience higher increase in college admissions, we may infer that less-developed provinces are less likely to retain college graduates though they experienced greater increases in college admissions.

Table 2.1 shows the logarithmic annual growth rate for the total number of employed workers and the number of college graduate workers by age group and workplace region. The total number of employed workers decreases by about 2% for those who were 30-39 years old in these 15 years. This is probably related to the family planning policies that began in late 1970s and early 1980s, which reduced the number of births. People born in this period started going into their 30s in 2001-2015, thus the number of employed workers is smaller than people the earlier cohort. Next, we look at the annual logarithmic change rate for employed college workers. The increase for college graduate workers is much

larger than that for all workers, which implies that skill upgrading is occurring over time in the labor market. Meanwhile, the number of college workers increased at a much higher rate for younger workers (20-29 year-olds) than for older workers, very likely due to the college expansion in China from 1999.

Panel B shows logarithmic annual change in number of employed workers and number of college graduate workers by workplace region. First, the growth rates for college graduates are much higher than those for all employed workers in all regions, indicating skill upgrading across China. The growth rates for both employed workers and college graduate workers are larger in the eastern provinces than in other regions. Since the annual increase in the number of students admitted to colleges in eastern, central, western and northeastern regions in 1992-2011 are 13.5%, 15.2%, 14.2% and 11.4% respectively, this indicates that some people who received their college education in the central and western provinces choose to work in the eastern provinces.

Next, we summarize the employed workers and workers with at least college degrees by gender, occupation, and industry. Due to the availability of these statistics in administrative source, we only list these statistics for urban residents in Table 2.2. For males, the growth rate for “clerk and business personnel” increased at a higher rate than for other occupations. For women, “manager and professional” workers increased in a higher rate than other occupations. The number of college graduate workers among both men and women increased at a much higher rate than the increase of all employment within each type of occupation, which further suggests an overall skill upgrading in the Chinese labor market. The increase in college workers among “farmer and production worker” is

larger than for other occupations, which is most likely because there were very small number of college workers doing farming or manufacturing work.

In panel B of Table 2.2, we see that the number of employed workers in the agricultural industry decreased by about 4.9% per year, while employment increased in the manufacturing and third service industries for both men and women. Joining the WTO in 2001 has greatly stimulated the growth in manufacturing industry, and further stimulated the growth in service industry due to the increase in personal income. For both men and women, the increase rate for college graduate workers in the second industry is much higher than in the agricultural and the service industries.

2.3 Data

In this chapter, we use the Chinese General Social Survey (CGSS) 2003, 2005, 2006, 2008, 2010, 2011, 2012, 2013, 2015 to study the impact of college expansion on the Chinese labor market. The CGSS samples were selected from urban areas and rural areas covering all provinces in mainland China. The sampling frames used for 2003-2006 CGSS, 2008 CGSS, and 2010-2015 CGSS are different.

The 2003–2006 CGSS sampling is based on a four-stage sampling design. It selected 125 counties²² in China, each covering 4 selected towns or city subdistricts. Then it selected 2 villages or neighborhood committees in each of the selected towns and selected 10 households in each of the selected villages. A person aged 18 and above was randomly selected from each sampled household to serve as respondent. There are 10,000 final

²² County unit in China is different from the county unit in the US. It is the third level of the administrative hierarchy in China, under the provincial level (the first level) and prefectural level (the second level, in China it is usually cities). Thus, a county is an administrative level under cities. It is most common in China that a city governs several counties.

respondents in the 2003-2006 CGSS samples. To select the 125 counties, they categorized Chinese counties under five categories (municipalities²³, provincial capital cities, other cities in the eastern region, central region, and western region). The CGSS 2008 differs from the CGSS 2003-2006 design in including the selection of provincial capital cities into the last three categories when selecting the sample counties. They cover a total of 6,000 final respondents. In contrast to the CGSS 2003-2008, the 2010-2015 CGSS employed a three-stage sampling frames, which omitted the second stage (selecting towns), and also treated Beijing, Shanghai, Tianjin, Guangzhou, and Shenzhen as self-representative stratum in the first stage (selecting counties). There are 12,000 final respondents in the CGSS 2010-2015 sample²⁴.

In the study of the college premium, our primary sample is of currently employed workers between 25-65 years of age who were living in urban areas at the time when they took the surveys. The total sample has 20,416 observations and the descriptive statistics for these observations are shown in Table 2.3 and Table 2.4.

In Table 2.3, we find that male respondents are under-represented in CGSS data. The Chinese Labor Statistical Yearbook shows that the average percentage of male workers is more than 60% in 2003-2015, as compared to about 55% male workers in the CGSS sample. The average age for the employed workers is 35 in the CGSS 2003, 37 in 2005-2008, and 40 in 2010-2015. The average age increases across years, which is similar to the statistics from Chinese Labor Statistical Yearbook (36 in 2003, 37 in 2007, and 38 in 2013).

²³ A municipality in China is a "city" with "provincial" power under a unified jurisdiction. As such it is simultaneously a city and a province of its own right (cited from Wikipedia). There are four municipalities in China: Beijing, Tianjin, Shanghai, Chongqing.

²⁴ For more detailed information on CGSS sampling design, see: <http://cgss.ruc.edu.cn/index.php?r=index/sample>

There is a large increase in the percentage of the college graduate workers, from 15% in 2003 to 30% in 2010-2015, which is a result of the college expansion policy. Annual income is the total annual income in the year previous to the survey (e.g. in CGSS 2003, income is for 2002), all adjusted to 2002 Chinese Yuan. We can see that the real income for urban workers increases over time. The percentage of people working in the eastern regions increases over time, while the percentage of people working in the central or western provinces declines. This is due to a migrant flow from central and western regions to eastern provinces during 2003-2015.

In Table 2.4, we compare the percentage of workers who are college graduates, the college premium, and the percentage of manager, professionals and technicians among workers with college degrees and those with less than college degrees, by gender and age in each of the three survey periods.

We can see from Table 2.4 that there was a large increase in percentage of college workers for both female and male young workers over time, but the increase for older workers was relatively smaller. For the wage premium, there was a large decrease for young female workers from 3.1 to 1.7 from 2002 to 2004-2007, and for young male workers, the premium decreased from 1.8 to 1.5. But for older workers, the decline in college premium was much smaller. Although the college premium for both male and female young workers increased from 2004-2007 to 2009-2014, it did not reach its original 2002 level. For older workers, their premium greatly increased from 2004-2007 to 2009-2014. Finally, the percentage of workers with at least college degrees of having high occupation (manager, professionals, and technicians) decreased for both young and older workers from 2003 to 2008, and it increased from 2008 to 2015. The percentage of workers

with less than college of having high occupation increased over time for older male workers, but it decreased over time for older female workers.

2.4 Empirical Models

Since the college expansion is mainly implemented through increasing the number of admissions to colleges, we use the number of students admitted to college as a measure of the college expansion. We then study impacts of the college expansion on the college premium, unemployment, and skills requirements in first jobs. Although the college expansion beginning in 1999 was a national policy, the implementation of this policy through increasing the college admissions differed across provinces (See Figure 2.2). Therefore, the impact of the college expansion can be identified by different level of increase in college admissions by province and across year.

The process of college admissions across provinces of China in our sampled period (1999-2014) can be described as followed: (1) The MOE announces a plan for total number of college admissions for the year, including total number of admissions for colleges under provincial government administration in each province and the number of admissions for each college under the central government's administration. These planning numbers are determined mainly based on the national five-year and ten-year plan for higher education, and also consider the national socioeconomic development in each year (Zhang, 2017). (2) Based on the planning number for each province by the MOE, the department of education in each province works with each provincial college to establish admission goals. (3) Each school determines the number of students to be admitted from each of the 34 provinces and then reports its plan back to the provincial department of education and publishes the plan

prior to the college entrance exam. In order to take account of the effect of economic development on admission planning, we control for provincial economic variables.

We use the following model to estimate the impact of the positive supply shock on several labor market outcomes:

$$y_{ipt} = \beta \widetilde{AD}_{p,t-4} + \boldsymbol{\gamma}' \mathbf{Edu}_i + \boldsymbol{\tau}' (\mathbf{Edu}_i * \widetilde{AD}_{p,t-4}) + \boldsymbol{\lambda}' \mathbf{X}_i + \boldsymbol{\varphi}' \mathbf{Z}_{p,t} + \delta_p + \delta_t + \varepsilon_{ipt} \quad (2.1)$$

y_i represents the outcome for a person i . We examine three outcomes in this chapter: log of annual income (adjusted to 2002 Yuan), a dummy for unemployment²⁵, the score of skill level for tasks used in the first job. $\widetilde{AD}_{p,t-4}$ is a centered measure of number of admitted student to college (unit: 0.1 million) four years before the year t when the person was surveyed ($t \in 2003, 2005, 2006, 2008, 2010, 2011, 2012, 2013, 2015$), where province p is the location of the residences of person i at the survey time. It was calculated as the number of students admitted to college in province p in $t-4$, subtracting the overall mean number of students admitted to college across the years 1998-2010 ($\widetilde{AD}_{p,t-4} = AD_{p,t-4} - \overline{AD}$), and then dividing by 100,000. \mathbf{Edu}_i is a vector dummies for educational level, $\mathbf{Edu}_i = [Edu_{i1}, Edu_{i2}, Edu_{i3}]$, including a dummy for 4-year college degree, a dummy for 2- or 3-year college degree, and a dummy for less than high school degree (the base group is high school degree). $\mathbf{Edu}_i * \widetilde{AD}_{p,t-4}$ is an interaction vector between personal educational attainment and college admissions in the province where the person lived in year $t-4$. \mathbf{X}_i is a vector for personal characteristics, including age, age squared and gender. $\mathbf{Z}_{p,t}$ is a vector for other time-variant provincial control variables, including provincial number of labor

²⁵ “Unemployed” equals 1 for those who did not have any job at the time of survey but had searched for a job in the three weeks prior to the survey.

force, log of provincial per capita GDP, per capita exports, per capita imports, fixed capital investment, percentage male, and percentage urban population in the province p in year t . We also include year fixed effects, δ_t , to control impacts of other national policies, and province fixed effects, δ_p , to control impacts of time-invariant differences among provinces. Since the impact of college expansion may differ by people's age and experience, we estimate the equation (2.1) separately for those who were 23-29, and those who were 30-65 in the year of survey.

To exclude the variation in the number of college admissions caused by economic growth, we use an alternative method to identify the impact of the college expansion policy on labor market outcomes. First, we regress the demeaned number of provincial college admissions in T ($\widetilde{AD}_{p,T}$) on the provincial GDP in $T-1$ and the provincial fixed effects; then, we use the residual from this regression ($\varepsilon_{p,T}$) as the proxy for the variation in the number of college admissions due to changes in provincial college admission policy that are not related to the economic growth, and use its values at year $t-4$ ($\varepsilon_{p,t-4}$, where t is the year of survey) to substitute the \widetilde{AD}_{t-4} in equation (2.1). Specifically,

$$\widetilde{AD}_{p,T} = \beta GDP_{p,T-1} + \delta_p + \delta_T + \varepsilon_{p,T} \quad (2.2)$$

$$y_{ipt} = \beta \varepsilon_{p,t-4} + \boldsymbol{\gamma}' \mathbf{Edu}_i + \boldsymbol{\tau}' (\mathbf{Edu}_i * \varepsilon_{p,t-4}) + \boldsymbol{\lambda}' \mathbf{X}_i + \boldsymbol{\varphi}' \mathbf{Z}_{p,t} + \delta_p + \delta_t + \varepsilon_{ipt} \quad (2.3)$$

where $\widetilde{AD}_{p,T}$ is a centered measure of number of provincial admitted student to college (unit: 0.1 million) in year T ; $GDP_{p,T-1}$ is the GDP in province p in year $T-1$; $\varepsilon_{p,T}$ is the residual from equation (2.2). Other variables have the same definitions as the variables in equation (2.1).

2.5 Main Results

2.5.1 Impact of the College Expansion on the College Premium

In this section, we estimate the impact on the college premium of the college expansion in 2002-2014 using equation (2.1). The CGSS survey solicits the respondent's annual income (including wages, bonuses, and other income) for the year prior to the survey. We thus have the income information for a total of 20,587 urban respondents from 31 provinces. We then combine the number of students admitted to colleges in each province and year with the CGSS data to see if the college expansion affects the college premium. The estimated results are shown in Table 2.5. In this analysis of the impact of the college expansion on the college premium, we must restrict the sample to urban residents since the question regarding income was only asked of them. We also limit the sample to respondents who were employed as full-time workers and have positive income.

Columns (1) – (4) show the results for younger workers (23-29 years old), who were more likely to be those just graduating from colleges and starting their careers. The estimated standard errors are clustered at the province level. However, since the number of provinces in China in our data is only 31, less than the minimum number of clusters that lead to reliable inference according to “rule of 42” (Angrist & Pischke, 2008), we apply the wild bootstrap method to construct reliable statistical tests for our estimated results. In this chapter, we use the wild cluster bootstrap test in the Stata (`boottest`) to determine the p-value for each of the estimated coefficients.²⁶

²⁶ It reports the t-statistic from Wald test and the p-value, which was generated from 999 wild cluster bootstrap samples that are assumed to follow the Rademacher distribution.

In column (1), we control province fixed effects and basic demographic variables (age, age squared, gender). The estimated college premium is 1.6 ($e^{0.476}$) at the average level of college admission for a province. The increase in college admissions decreases the college premium. As the number of admitted students in a province increases by 0.1 million, the college premium for young workers decreases by about 8.6 log points. To rule out the possible effects from changes in the overall labor force, we include 15-64 population in each province in column (2). The results change little. In column (3), we add several provincial economic and demographic variables to the model in order to control for changes in other socioeconomic trends in each province, including log annual GDP, log investment, log imports and log exports, percentage male of the population, percentage of population living in urban area in the province. Both the college premium and the negative impact of the college graduates decrease slightly, but not in a significant way. In column (4), we further control year fixed effects, and it does not change the results.

In Figure 2.10 we plot the estimated college premium for young workers in 2002-2014 from a simple model which controls only personal characteristics (age, age squared, gender), province fixed effects, and time fixed effects (the line with circle markers), and compare it with the predicted college premium using coefficients from column (4) in Table 2.5 (the line with square markers), and with the predicted college premium for young workers using equation (2.1) without including college admissions (the line with triangle markers). We also plot the total national college admission from 1994 to 2010 (unit: million, the line with cross markers, the numbers shown on the graph are for calendar year $T+4$).

To obtain the predicted college premiums using equation (2.1), we first predict the income in 2002-2014 for each person between ages 23-29. We substitute in estimated coefficients (using the reported coefficients in column (4) of Table 2.5), personal characteristics, provincial economic and demographical measures, and the number of students admitted to college in each province and year into equation (2.1). Then we average the predicted income for all college graduates and for all high school graduates to calculate the average college premium in each survey year.

In Figure 2.10, we observe decreases in college premium in 2002-2007, 2009-2012 and an increase in 2007-2009 and 2012-2014 for the estimated college premiums using the simple model. The predicted college premiums based on the estimated coefficients in column (4) of Table 2.5 show continuous decreases from 2002-2014. And the predicted college premiums using the model without including college admissions show a relatively flat line in contrast to both the estimates from the simple model and model including college admissions. That the predicted college premium declines even when the effects of college admission are excluded suggests that the positive supply shock due to college expansion only explains part of the decrease in college premiums, and there may be demand shifters which further decreased the college premiums. However, our model including college admissions still predicts the decline of the college premiums better than the model without college admissions.

The impact of college expansion on older workers (30-65 years old) is different. The first rows of Columns (5) - (8) report the estimated college premium for older workers is about 80%, which is greater than the college premium for younger workers. Also, we do not observe a decline in the college premium due to the increase in college admission for

older workers. In contrast, after controlling for the population age 15-64, we found that the college expansion increased the college premium by about 4 percent for older people who completed college degrees. A possible explanation for the different impact of college expansion on young and older skilled workers is that older skilled workers have more experience, thus new college graduates cannot easily substitute for them. Moreover, experienced skilled workers can help training the increasing numbers of young skilled workers, causing their incomes to increase in response to the college expansion.

The reduction in the college premium due to college expansion may result from changes of the college premium within occupation and industry, or movement between occupation and industry, or both. Therefore, we control for the industry and occupation fixed effects in equation (2.1), in order to identify the changes of college premiums within industry and occupation.

Unfortunately, not every year of CGSS surveys has information on the current industry and occupation of employment. To make our results comparable, we use the CGSS 2003, 2006, 2010, 2011, 2012, 2013 (9,463 observations) to conduct this estimation, since these surveys contain full information of the occupation and industry a respondent works in. The survey codes 15 industries and 37 occupations.

The estimated results controlling for industry and occupation fixed effects are shown in APTable 2.1. In columns (1) (for young workers) and (5) (for older workers) of APTable 2.1, we do not control industry or occupation fixed effects. In columns (2) and (6), we control for industry fixed effects and in columns (3) and (7), we control for occupation

fixed effects. Finally, in columns (4) and (8), we control for both industry and occupation fixed effects.

As can be seen from the table, comparing with estimates when no industry or occupation fixed effects have been controlled (the college premium decreases by 20.3 log percentage points), the decline of college premium due to college graduates within the same industry for younger workers due to the increase in college admission is slightly smaller (decreases by 20 log percentage points). When controlling for occupation, we also observe a decrease in college premium by 19.6 log percentage points per 0.1 million increase in college admission, which is slightly smaller compared with column (1). When controlling for both industry and occupation fixed effects, we observe a 19.8 log percentage points decrease in college premium as the college admissions increase by 0.1 million. These results suggest the college admission decreased the college premium within industry and occupation, and it also decreased the college premium between industry and occupation since the estimated impacts are greater when not controlling the industry and occupation fixed effects. We do not find a significant decline in the college premium for older workers due to increase in college admission in any specification.

To exclude the variation in college admissions caused by GDP, we use equations (2.2) and (2.3) to re-estimate the effect of the college expansion policy on income. The results are listed in columns (1) and (3) of APTable 2.2. The results using the alternative method confirm the patterns observed from Table 2.5, which suggest that as the number of college admissions increases, the college premium decreases for young workers, and increases for older workers.

Some may concern about the issue that the scale of our measure using raw number of college admissions is not well controlled. In equation (2.1), we control for the provincial population as well as the province fixed effects to control the possible factors that may affect the scale of provincial college admissions. In APTable 2.3, we also use the ratio between the number of provincial college admissions and population in year t as the measure of the college expansion to conduct the robustness check. The patterns we observe from the table are consistent with our results in Table 2.5.

We also conduct a placebo test in order to check whether the number of provincial college admissions in year $t-4$ affects the other macro-level variables in year t , which could potentially bias our results. In APTable 2.4, we list the regression results of provincial population, GDP, and unemployment rate on the number of provincial college admissions in year $t-4$, controlling for the lags of the outcome variables from $t-1$ to $t-4$. As can be seen from the results, the number of provincial college admissions in year $t-4$ has insignificant impact on other current macro-level variables, and thus we may exclude the possible bias that the number of college admissions affect individual's outcome through other macro-level variables.

In summary, we concluded that the impact of college expansion on the college premium differs for young and older workers. It drives down the college premium among young workers while it has no significant negative impact on the college premium on older workers. We also observe a decline in the college premium for young workers within industry and occupation. This is consistent with the view that when there is a positive supply shock of new skilled labor, young skilled workers are more likely than older workers to suffer a substitution effect that reduces their relative incomes. Our results are

also consistent with Freeman (1977), who found a decline in relative wages for young college educated workers during the 1970s due to a large increase of college educated labor in the U.S.

We also use the 2SLS method to identify the impact of a positive shock brought from the college expansion on people's income in appendix section A1.

2.5.2 Impact of the College Expansion on Unemployment

In the previous subsection, we studied the impact of college expansion on the college premium among younger and older workers. In the short run, college expansion may also create a surplus in skilled labor if companies do not increase their hiring of skilled workers sufficiently. Also, if the quality of college education decreases, the difference between the marginal product produced by college graduates and people with high school degrees may shrink, which may result in a decreased demand of college graduates. Increasing unemployment of college graduates is a phenomenon which has aroused wide concern in China, with popular media accounts and magazines attributing the unemployment of college graduates to the college expansion beginning in 1999.²⁷

Based on the Labor Statistics Yearbook (1997-2015), Figure 2.11 plots the percentage of college graduates among urban unemployed workers (the solid-square line) and the percentage of college graduates among urban employed workers (the hollow-square line); as well as the percentage of college graduates among 23-29-year-olds urban unemployed workers (the solid-circle line) and percentage of college graduates among urban 23-29-year-olds unemployed workers (the hollow-circle line). As can be seen from the graph, the

²⁷ See Pan (2005), Chen (2006), Xu (2007), Dong (2012), and Zhang (2013).

percentage of college graduates increases among both employed and unemployed workers, suggesting a trend of overall skill upgrading in the labor force. Before 2003, the percentage of college graduates among employed and unemployed workers are about the same for both young workers and for all workers. However, the percentage of college graduates among unemployed workers starts to exceed the percentage of college graduates among employed workers after 2003, which is the graduation year for those admitted to 4-year colleges in 1999, which is the beginning year of the college expansion. The gap continues to enlarge in the following years. Moreover, the gap in the percentage of college graduates between unemployed and employed workers among young people is greater than the gap between all unemployed and employed workers. This observation indicates that the sudden increase in skilled labor makes it harder for college graduates, especially for young college graduates, to find jobs.

In this section, we use our model to explore the impact of college expansion on the likelihood of unemployment for college graduates. The dependent variable of equation (2.1) in this section, $Unemp_{i,p,t}$, is a dummy for the unemployment status for a person i living in province p in year t . The model thus becomes a linear probability model (LPM). In the regression, we only include people who were not in school at the time of the survey and exclude those who were out of the labor force (i.e., not seeking jobs).

Table 2.6 reports the results of unemployment model described above by age group. Alternative specifications are the same as in Table 2.5. The available sample to conduct this regression has 30,072 observations. Based on the table, we first observe that both 4-year and 2- or 3-year college graduates are less likely to be unemployed than workers with high school degrees or below. The probability of unemployment for a young 4-year college

graduate is about 6-7 percentage points less than the probability for a high school graduate, and the probability for a 2- or 3-year college graduate about 7-8 percentage points less than the probability for a high school graduate, which is less than a 4-year college graduate. For older college graduates, their relative advantage in terms of employment is greater than the advantage for young college graduates. Older 4-year college graduates are 8-10 percentage points less likely to be unemployed than high school graduates, and older 2- or 3-year college graduates are 10 percentage points less likely to be unemployed than high school graduates.

The college expansion has significantly increased the probability of unemployment for 4-year and 2- or 3-year college graduates in both age groups. For young 4-year college graduates, a 0.1 million increase in provincial college admission increases their probability of unemployment by 3-4 percentage points relative to people with high school degrees (the difference in probability of unemployment between 4-year college graduates and high school graduates decreases by about 58%). Our estimates vary little across different model specifications.

The impact is smaller for older 4-year college graduates. It increases their relative probability of unemployment by 2-3 percentage points (the difference in the unemployment probability between 4-year college graduates and high school graduates decreases by about 22%). The college expansion also increased the probability of unemployment for 2- or 3-year college graduates. For every 0.1 million increase in college admission, the probability of unemployment for a young 2- or 3-year college graduate increased by about 5 percentage points, and it increases the probability for older college graduates by about 3 percentage points. The results suggest that college expansion increased the probability of

unemployment for both young and older skilled people. Since we do not observe any significant negative impact of the college expansion on the college premium for older workers, the college expansion affects for older skilled workers mainly work through increasing their probability of unemployment. However, the negative impact in terms of unemployment is greater for young college graduates than older college graduates.

We also observe a decrease in the probability of unemployment for people with less than a high school education for both young and older workers in response to the college expansion. Young people are more likely to pursue a college degree in the college expansion period since the likelihood of enrolling in college for them is much greater, as the college premium is still large. Thus, a smaller proportion of young people choose to end up with high school degrees or below, which may create a shortage of low-skilled workers and thus increase the likelihood of employment for people with low skills.

In columns (2) and (4) of APTable 2.2, we re-estimate the effect of the college expansion policy on unemployment based on equations (2.2) and (2.3) to exclude the variation in college admissions caused by other macro-level factors. The results also confirm the patterns we observed in Table 2.6.

We use the ratio of the number of provincial college admissions and the provincial population as the measure for the college expansion, and list its regression results on the probability of unemployment in columns (2) and (4) of APTable 2.3. The results are consistent with Table 2.6.

Instead of using the linear probability model, we assume the logistic probability. APTable 2.5 reports the estimated marginal effects using a Logit model applying to

equation (2.1). Similar to the results assuming linear probability (Table 2.6), we observe that the college graduates (both 2- or 3-year and 4-year) are less likely to become unemployed than high school graduates, while people with less than high school degree education are in general more likely to be unemployed. The college expansion significantly increased the probability of unemployment for young college graduates. The probability increased about 4 percentage points for both young 4-year college graduates and young 2- or 3-year college graduates. The impact of college expansion on the probability of unemployment is not statistically significant for older workers. We also observe a decrease in unemployment likelihood for both young and old people with less than high school degrees. Although the estimates using the logistic probability model slightly differs in magnitude from those of the linear probability model in Table 2.6, the pattern is essentially the same: we observe an increase in the probability of unemployment for both young and older college worker as the college admission increases, and the impacts for older college graduates are either small or not statistically significant.

In summary, our results suggest the college expansion has significantly reduced the advantage in terms of the likelihood of employment for both 4-year and 2- or 3-year college graduates relative to people with high school degrees or less education. The negative impact due to college expansion is greater for younger people than for older people in terms of the percentage decrease in the relative employment advantage over high school graduates. These results further support our suggestion that young college graduates are more easily substituted for young college graduates than for older college graduates. We therefore conclude that the positive supply shock of young skilled labor affects not only

the relative income between skilled and unskilled workers but it also their relative probability of employment.

Similar to our analyses of income, we use the 2SLS method to identify the impact of a positive shock due to the college expansion on people's unemployment in appendix section A1 and ATable 2.9.

2.5.3 Impact of the College Expansion on Tasks of First Job

In this section, we study the impact of the college expansion on the skills used in the first jobs individuals obtain, which provides another perspective of how the college expansion affects young workers. We follow previous studies that have considered changes in cognitive and manual tasks required in jobs.

A job that requires intensive cognitive skill usage is more likely to be filled by college workers and is compensated with higher income since college graduates have a comparative advantage in dealing with cognitive tasks, and a job which has more manual tasks is more likely to be filled by noncollege workers since they have a comparative advantage in dealing with manual tasks. When an increased number of college graduates enter the job market, it may be harder for them to find jobs that use their cognitive ability. A decrease in quality of college education due to the increase in college admission also contributes to a reduction in skills learned from colleges, which may therefore decrease the demand for new college graduates. As a result, new college graduates may end up taking jobs that use less cognitive skills.

We use the 2003, 2006, 2008 CGSS data to study the impact of college expansion on the cognitive and manual task scores for the first job, using the method described in

Almeida (2018). The CGSS survey for these years²⁸ provides the information of ISCO-88 codes for respondents' occupations. We convert the occupation codes to the U.S. Standard Occupational Classification (SOC) 2010 using a crosswalk provided by the Institute for Structural Research (IBS)²⁹. After the conversion process, we are able to merge the CGSS data with the O*NET evaluation data for each occupation. Although there may be differences within occupation between China and US or Europe, we assume that the classification still preserves the order of different skills usages.

In the evaluation data, there are “importance scores” from 1 (least important) and 5 (most important) for 52 abilities associated with an occupation. Since, in the O*NET data, the importance scores for tasks in an occupation do not change across time, the variation in scores is due to people having different occupations. By connecting CGSS to O*NET, we assume occupations in China have similar activities or at least a similar ranking of tasks as the same occupations in the U.S.

Following Almeida (2018), we group the tasks in each job into four categories: manual precision tasks, other manual tasks, cognitive analytical tasks, and cognitive-communication tasks. We average the importance scores within each category of tasks for every job. We assume that higher average scores for a task in an occupation imply that people within an occupation with particular tasks are more likely to apply corresponding skills in their work.

²⁸ Although CGSS 2005 has occupational information for the first job, it is not coded using the ISCO-88 standard.

²⁹ See <http://ibs.org.pl/en/resources/occupation-classifications-crosswalks-from-onet-soc-to-isco/> for more detail.

Normally, people with college degrees are more likely to have jobs that require cognitive skills and less likely to have jobs requiring high scores on manual tasks. However, a large positive shock in the supply of skilled labor may temporarily make it hard for college graduates to find jobs that fully employ their cognitive skills. If the vacancies of jobs that require lower levels of cognitive skills are not yet be filled, college graduates may end up with these jobs. Since the return to cognitive skill is on average greater than the return to manual skills, this mechanism provides another explanation of the reduction in the college premium. Whereas the CGSS does not have information for earnings on first jobs, it asked people the detailed types for their first jobs. In the previous section, we have shown that the effect of college expansion on income and the probability of unemployment is greater for young college graduates than older college graduates. Therefore, in this section, we narrow our analysis on people's first jobs, which were most likely obtained at their earlier age and after finishing their highest education.

Whereas the CGSS does not have information for earnings on first jobs, it asked people to provide details regarding their job activities. In the previous section, we showed that the effect of college expansion on income and the probability of unemployment is greater for young college graduates than older college graduates. In this section, we narrow our analysis on people's first jobs, which were most likely obtained after finishing their highest education.

We use $Tscore_{i,j,p,t}$ as the dependent variable in equation (2.1) denoting the average scores for a task group j of the first job for respondent i who works in province p in year t . We link the people's first occupation with O*NET by occupation codes, exclude respondents who received their college degrees after the year when they started their first

job (211 observations), and the total number of observations for this analysis is 1,705. Table 2.7 reports the estimation results for equation (2.1). In contrast to prior estimation, which used the income or employment at the time of the survey, job information is for the individual's first job.

As can be seen from the table, the first jobs for 4-year and 2- or 3-year college degree graduates demand higher levels of cognitive skills and lower levels of manual skills than jobs of those with high school degrees or below. For example, the average cognitive analytical score of first jobs for 4-year college graduates is 0.05 more than those for people in the same provinces who have high school degrees, which is approximate 11 times of its standard error.

The college expansion decreases the cognitive scores of the first jobs for 4-year college graduates relative to high school graduates. As the number of students admitted to college increase by 0.1 million in each province, the difference of analytical scores between 4-year college graduates and high school graduates decreases by 0.02 (3 times of its standard error), and the difference of communicative score goes down by 0.05 (1.8 times of its standard error). Although the relative scores of manual precision and other manual tasks for college graduates increase as college admission increases, the results are not statistically significant. Another interesting finding is that the college expansion decreases the average manual-precision scores of first jobs for people with high school degrees. As the number of college admission increase by 0.1 million, the manual-precision scores for the first jobs of high school graduates decrease by 0.07 (3 times of its standard error).

In summary, with increased college admissions, college graduates are losing their advantage over high school graduates in landing jobs requiring higher cognitive skills due to the college expansion. We have identified a possible channel of how the college expansion may reduce the college premium. It reduces the chance that a college graduate to finds a job requiring greater cognitive skills, which therefore decreases average income relative to a high school graduate. In APTable 2.6, we estimate the return to cognitive and manual scores among full-time workers, controlling for their educational level, gender, age, age squared, province fixed effects, and year fixed effects. As can be seen from the table, jobs with higher cognitive scores have a higher average income. In the short term, the vacancies for higher-paid jobs are limited, so some college graduates are unable to land these jobs. Instead, they may have to accept lower-paid jobs that require lower levels of cognitive skill. This implies that the college premium is reduced in part due to the changes in the occupational distribution for college graduates.

2.6 Conclusion

The college expansion that began in 1999 has significantly enlarged the skilled labor force in China within a short period. It thus created a positive supply shock of skilled labor which substantially changed labor market behaviors and outcomes. In this chapter, we explore short run labor market responses to the college expansion, as well as the market responses to the positive supply shock brought about by the college expansion. We estimate the impact of college expansion on the college premium, unemployment, and skills used in first jobs. We also use the number of students admitted to college as an instrument of supply shock and study its impacts on the college premium and unemployment.

Our results suggest that the college expansion has significantly reduced the college premium for young college graduates (age 23-29) in the short run. A 0.1 million increase in all college admissions (standard colleges and adult colleges) reduces the log college premium for young people by 9 log percentage points. However, we do not observe significant impacts of college expansion on the college premium for older workers (age 30-65). Consistent with other studies, people who just graduated from colleges are more likely to suffer a substitution effect than older workers. We also identify a decrease in the college premium within industry and occupation among young workers due to the college expansion.

College graduates are also more likely to be unemployed in the short run as the college admissions increase. Our results suggest that the college expansion significantly increases the probability of unemployment for both young and old college graduates. The impact of college expansion on probability for young college graduates is greater than older college graduates, which further supports the substitution effect hypothesis.

The college expansion also reduces the likelihood of college graduates finding their first jobs which require intensive use of cognitive skills. Therefore, the college expansion not only decreased the college premium within occupation or industry, but also implies that new college graduates are more likely to accept occupations requiring lower levels of cognitive skills and lower average salaries.

The results in chapter suggest that a positive supply shock on skilled labor bring a negative impact to younger college graduates but has less negative impact on older college graduates in the short term.

Table 2.1 Annual Increase Rate for Number of Employed Workers and Number of College Graduate Workers

	Total Employed Workers (Annual log change *100)	Employed College-Educated Workers (Annual log change *100)
Panel A		
By Age Group		
<i>20 – 24</i>	0.60	12.20
<i>25 - 29</i>	1.11	11.36
<i>30 – 34</i>	-1.74	8.59
<i>35 – 39</i>	-2.73	6.31
<i>40 – 44</i>	2.44	8.71
<i>45 – 49</i>	2.41	9.02
<i>50 – 54</i>	-0.07	7.94
<i>55 – 59</i>	2.88	7.04
Panel B		
By Region		
<i>Eastern</i>	2.30	10.98
<i>Central</i>	0.29	10.01
<i>West</i>	-6.00	9.98
<i>Northeastern</i>	-3.00	7.01

Note: This table reports the logarithmic annual growth rate in 1996-2015 for the total number of employed workers and the number of college graduate workers by age group and work region. The logarithmic annual growth rate is estimated by regressing log number of employed workers or workers with college degrees on a linear time trend.

Table 2.2 Annual Increase Rate for Number of Employed Workers and Number of College Graduate Workers (Urban)

	Total Employed Workers (Annual log change *100)		Employed College Graduate Workers (Annual log change *100)	
	Male	Female	Male	Female
Panel A				
By Occupation				
<i>Manager, Professional</i>	4.34	7.19	6.58	9.57
<i>Clerk, Business Personnel</i>	6.71	5.42	10.43	10.13
<i>Farmer, Production Worker</i>	4.90	1.63	18.64	15.02
Panel B				
By Industry				
<i>Agricultural</i>	-4.84	-4.92	0.76	2.83
<i>Manufacturing</i>	6.20	4.00	12.30	12.22
<i>Service</i>	3.85	4.51	5.27	7.31

Note: This table reports the logarithmic annual growth rate in 1996-2015 for the total number of employed workers and the number of college graduate workers in urban areas by occupation and industry. The logarithmic annual growth rate is estimated by regressing log number of employed workers or workers with college degrees on a linear time trend.

Table 2.3 Descriptive Statistics of Sample in CGSS 2003-2015 (Urban)

	CGSS 2003 (N=2,926)		2005-2008 (N=8,116)		2010-2015 (N=9,374)	
	Sample Mean (SD)	Census 2000 Mean	Sample Mean (SD)	Census 2010 Mean	Sample Mean (SD)	Census 2010 Mean
Male	0.49 (0.50)	0.51	0.47 (0.45)	0.51	0.50 (0.50)	0.51
Age	38.99 (14.33)	32.37	40.77 (13.27)	35.37	41.65 (14.24)	35.37
College	0.12 (0.32)	0.12	0.16 (0.33)	0.17	0.24 (0.33)	0.17
Region						
<i>Eastern</i>	0.38	0.42	0.41	0.45	0.49	0.45
<i>Central</i>	0.23	0.24	0.22	0.23	0.19	0.23
<i>Western</i>	0.28	0.26	0.25	0.22	0.18	0.22
<i>Northeastern</i>	0.11	0.09	0.12	0.09	0.14	0.09
Annual income (¥)	10,592 (13,976)	10,834	18,776 (17,620)	20,021	42,362 (50,902)	45,383

Note: This table reports the mean and standard deviations of basic demographic information, including male, age, college degree attainment, annual income, proportion of resident region for urban sample in the CGSS 2003-2015 by survey period.

Table 2.4 Statistics for Labor Market Variables in CGSS 2003-2015 (Urban)

	CGSS 2003 (N=2,926)		2005-2008 (N=8,116)		2010-2015 (N=9,374)	
	Male	Female	Male	Female	Male	Female
Fraction of College Workers						
20-29	0.18 (0.55)	0.19 (0.50)	0.34 (0.42)	0.35 (0.42)	0.58 (0.53)	0.57 (0.48)
30-65	0.15 (0.33)	0.12 (0.28)	0.20 (0.35)	0.17 (0.33)	0.23 (0.41)	0.18 (0.37)
College Premium	Income in 2002		Income in 2004-2007		Income in 2009-2014	
20-29	1.81	3.13	1.52	1.70	1.66	2.21
30-65	1.81	2.14	1.60	2.14	2.41	3.06
% of Manager, Professional, Technician						
(i) College & above						
20-29	0.60 (0.48)	0.51 (0.47)	0.50 (0.40)	0.45 (0.40)	0.58 (0.55)	0.62 (0.48)
30-65	0.68 (0.33)	0.69 (0.30)	0.49 (0.43)	0.53 (0.41)	0.64 (0.47)	0.66 (0.44)
(ii) High school & below						
20-29	0.06 (0.38)	0.20 (0.55)	0.21 (0.38)	0.18 (0.36)	0.19 (0.40)	0.20 (0.39)
30-65	0.10 (0.29)	0.24 (0.38)	0.15 (0.32)	0.15 (0.32)	0.15 (0.35)	0.13 (0.33)

Note: This table reports the mean and standard deviations for fraction of college graduates among employed workers, college premium, and percentage of manager, professional or technician among people with college degrees and people with less than college degrees. The standard deviations are shown in the parentheses.

Table 2.5 Regression Results of the College Expansion on Annual Income

	Age in 23-29				Age in 30-65			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Col	0.476*** (0.036)	0.486*** (0.040)	0.438*** (0.046)	0.435*** (0.040)	0.628*** (0.062)	0.598*** (0.027)	0.569*** (0.026)	0.574*** (0.026)
Lcol	0.172*** (0.036)	0.190*** (0.037)	0.149*** (0.035)	0.154*** (0.035)	0.349*** (0.032)	0.343*** (0.020)	0.335*** (0.020)	0.334*** (0.020)
Lhs	-0.093* (0.048)	-0.094** (0.047)	-0.158*** (0.051)	-0.153*** (0.040)	-0.260*** (0.028)	-0.268*** (0.016)	-0.286*** (0.016)	-0.287*** (0.016)
\widetilde{AD}_{t-4}	0.262*** (0.033)	0.255*** (0.067)	-0.002 (0.043)	-0.065 (0.039)	0.253*** (0.024)	0.191*** (0.014)	-0.014 (0.020)	-0.035 (0.025)
Col * \widetilde{AD}_{t-4}	-0.086*** (0.016)	-0.096*** (0.024)	-0.088*** (0.026)	-0.088*** (0.029)	-0.009 (0.026)	0.043** (0.019)	0.040** (0.019)	0.038** (0.019)
Lcol * \widetilde{AD}_{t-4}	-0.005 (0.016)	-0.022 (0.022)	-0.020 (0.021)	-0.019 (0.027)	-0.010 (0.026)	0.012 (0.015)	0.003 (0.015)	0.006 (0.015)
Lhs * \widetilde{AD}_{t-4}	-0.050** (0.022)	-0.001 (0.040)	0.006 (0.037)	-0.003 (0.029)	-0.042*** (0.013)	-0.007 (0.012)	-0.005 (0.012)	-0.005 (0.012)
X_i	Yes							
δ_p	Yes							
\widetilde{TL}	No	Yes	Yes	Yes	No	Yes	Yes	Yes
$Z_{p,t}$	No	No	Yes	Yes	No	No	Yes	Yes
δ_t	No	No	No	Yes	No	No	No	Yes
N	3,735	3,735	3,735	3,735	16,852	16,852	16,852	16,852
R ²	0.3364	0.3385	0.3710	0.3838	0.3863	0.3879	0.4054	0.4090

Note: This table reports the regression results for equation (2.1). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is the log of annual income for full-time urban employed respondents. \widetilde{AD}_{t-4} is the provincial college admission (both 2- or 3-year and 4-year) in $t-4$, centering around its mean level across all provinces and all years. *Col*, *Lcol*, *Lhs* are dummies for people with 4-year college degrees, people with 3-year college degrees, and people with less than high school education, respectively. The regression excludes people with more than 4-year college education. Columns (1)-(4) list estimates for young workers (23-29-year-olds), and columns (5)-(8) list estimates for older workers (30-65-year-olds). We control personal demographic variables X_i (age, age squared and gender), and province fixed effects δ_p in columns (1) and (5), controlling changes in overall labor force and its interaction with each education dummies \widetilde{TL} in columns (2) and (6), and adding provincial economic and demographic variables $Z_{p,t}$ (log per capita GDP, fixed capital investment, import, export, percentage of male and percentage of urban population) in columns (3) and (7), and including year fixed effects δ_t in columns (4) and (8). Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

Table 2.6 Regression Results of the College Expansion on Unemployment

	Age in 23-29				Age in 30-65			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Col	-0.066*** (0.012)	-0.079*** (0.012)	-0.070*** (0.014)	-0.063*** (0.014)	-0.102*** (0.009)	-0.109*** (0.009)	-0.090*** (0.009)	-0.081*** (0.009)
Lcol	-0.075*** (0.014)	-0.086*** (0.011)	-0.082*** (0.012)	-0.081*** (0.012)	-0.100*** (0.009)	-0.101*** (0.010)	-0.097*** (0.010)	-0.102*** (0.009)
Lhs	0.078*** (0.021)	0.075*** (0.020)	0.078*** (0.020)	0.076*** (0.020)	0.081*** (0.009)	0.084*** (0.009)	0.083*** (0.008)	0.080*** (0.008)
\widetilde{AD}_{t-4}	-0.051*** (0.008)	-0.067*** (0.016)	-0.028 (0.019)	-0.011 (0.019)	-0.078*** (0.013)	-0.066*** (0.020)	0.017 (0.016)	0.015 (0.011)
Col * \widetilde{AD}_{t-4}	0.025*** (0.006)	0.042*** (0.009)	0.041*** (0.010)	0.037*** (0.010)	0.026*** (0.004)	0.031*** (0.006)	0.025*** (0.005)	0.018*** (0.006)
Lcol * \widetilde{AD}_{t-4}	0.026*** (0.008)	0.049*** (0.009)	0.049*** (0.009)	0.049*** (0.009)	0.026*** (0.005)	0.030*** (0.008)	0.032*** (0.007)	0.035*** (0.007)
Lhs * \widetilde{AD}_{t-4}	-0.025** (0.010)	-0.049*** (0.016)	-0.049*** (0.016)	-0.049*** (0.016)	-0.029*** (0.004)	-0.044*** (0.007)	-0.041*** (0.007)	-0.043*** (0.007)
X_i	Yes							
δ_p	Yes							
\widetilde{TL}	No	Yes	Yes	Yes	No	Yes	Yes	Yes
$Z_{p,t}$	No	No	Yes	Yes	No	No	Yes	Yes
δ_t	No	No	No	Yes	No	No	No	Yes
N	5,334	5,334	5,334	5,334	24,738	24,738	24,738	24,738
R ²	0.0702	0.0787	0.0843	0.0954	0.1116	0.1129	0.1337	0.1537

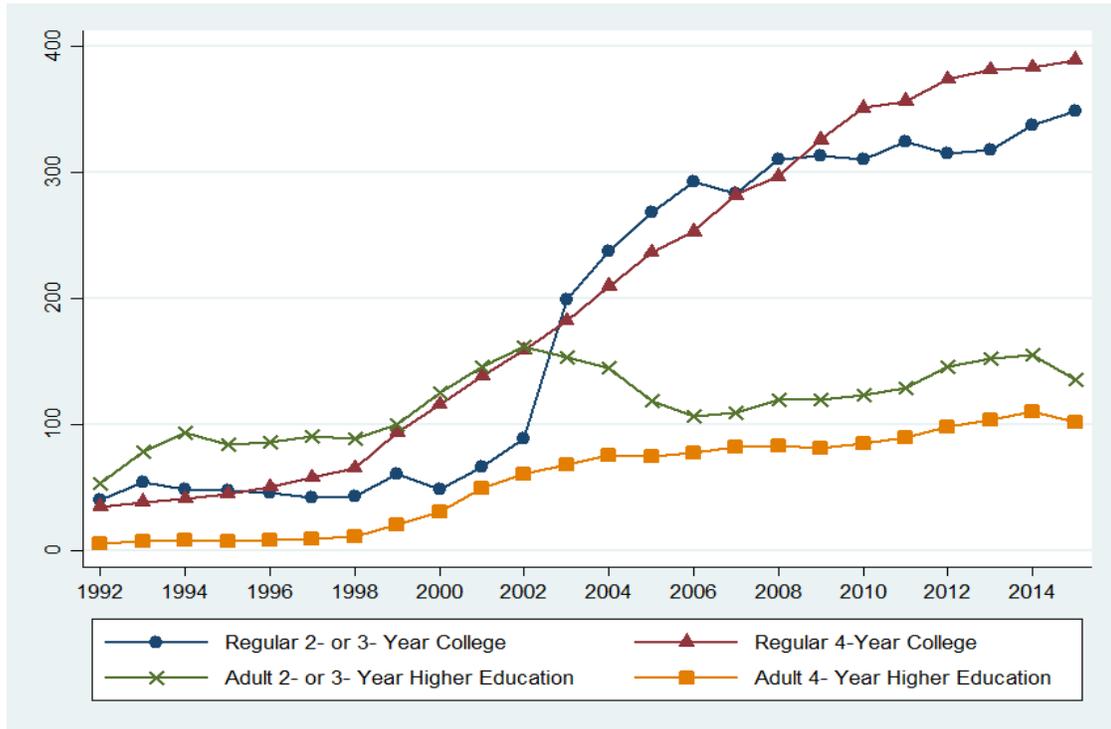
Note: This table reports the regression results for equation (2.1). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is the dummy for unemployment for urban respondents. \widetilde{AD}_{t-4} is the provincial college admission (both 2- or 3-year and 4-year) in $t-4$, centering around its mean level across all provinces and all years. *Col*, *Lcol*, *Lhs* are dummies for people with 4-year college degrees, people with 3-year college degrees, and people with less than high school education, respectively. The regression excludes people with more than 4-year college education. Columns (1)-(4) list estimates for young workers (23-29-year-olds), and columns (5)-(8) list estimates for older workers (30-65-year-olds). We control personal demographic variables X_i (age, age squared and gender), and province fixed effects δ_p in columns (1) and (5), controlling changes in overall labor force and its interaction with each education dummies \widetilde{TL} in columns (2) and (6), and adding provincial economic and demographic variables $Z_{p,t}$ (log per capita GDP, fixed capital investment, import, export, percentage of male and percentage of urban population) in columns (3) and (7), and including year fixed effects δ_t in columns (4) and (8). Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

Table 2.7 Regression Results of the College Expansion on Task Scores of First Job

	Cognitive Analytical	Cognitive Communicative	Manual Precision	Manual Others
Col	0.045*** (0.004)	0.294*** (0.019)	-0.181*** (0.015)	-0.063*** (0.005)
Lcol	0.032*** (0.004)	0.225*** (0.027)	-0.141*** (0.018)	-0.054*** (0.005)
Lhs	-0.035*** (0.006)	-0.216*** (0.031)	0.116*** (0.020)	0.034*** (0.006)
\widetilde{AD}_{t-4}	0.011 (0.007)	0.040 (0.031)	-0.066*** (0.022)	-0.009 (0.006)
Col * \widetilde{AD}_{t-4}	-0.015*** (0.005)	-0.051* (0.028)	0.018 (0.022)	-0.010 (0.006)
Lcol * \widetilde{AD}_{t-4}	-0.007 (0.006)	-0.018 (0.026)	0.019 (0.021)	-0.002 (0.006)
Lhs * \widetilde{AD}_{t-4}	0.004 (0.008)	0.003 (0.032)	0.014 (0.028)	-0.001 (0.006)
X_i	Yes	Yes	Yes	Yes
δ_p	Yes	Yes	Yes	Yes
\widetilde{TL}	Yes	Yes	Yes	Yes
$Z_{p,t}$	Yes	Yes	Yes	Yes
δ_t	Yes	Yes	Yes	Yes
Mean	0.452	3.447	1.212	0.471
SD	0.072	0.367	0.251	0.075
N	1,705	1,705	1,705	1,705
R ²	0.2203	0.2758	0.2038	0.2318

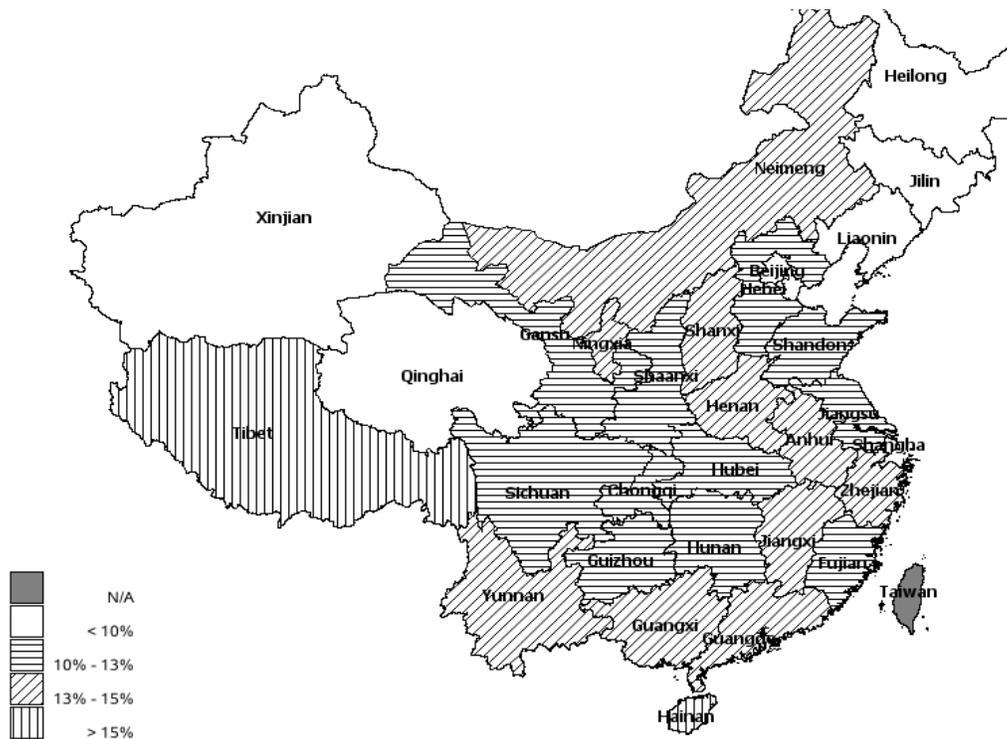
Note: This table reports the regression results for equation (2.1). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is the average score of a skill j used in the first job of the respondent in 1998-2008. We exclude respondents who received their college degree after the year when they started their first job. All of the variables listed in the table have the same definition as shown in Table 2.5 and Table 2.6. To conduct this analysis, we use the CGSS 2003, 2006, 2008. We control the provincial economic variables, and personal demographic variables, province fixed effects, changes in overall labor force and its interaction with each education dummies in province p in year t , and year fixed effects. Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

Figure 2.1 Number of Students Admitted to Regular and Adult 4-Year and 2- or 3-Year Higher Education Institutions in 1992-2015



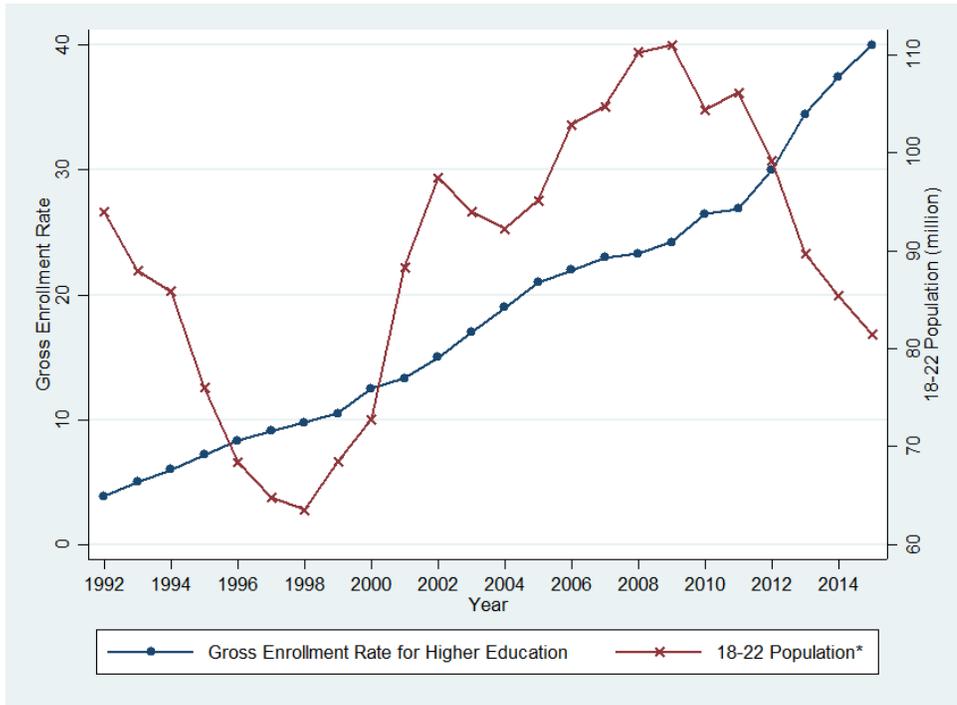
Note: This figure plots the number of students admitted to regular and adult higher education institutions in 1992-2015 (Unit: 1,000). The line with round dot denotes the number of students admitted to 2- or 3-year colleges, the line with triangle dot denotes the number of students admitted to 4-year colleges, the line with cross denotes the number of students admitted to 2- or 3-year adult colleges, and the line with square denotes the number of students admitted to 4-year adult colleges (Source: Educational Statistics Yearbook of China).

Figure 2.4 Average Annual Increase Rate of College Admissions in 1998-2012



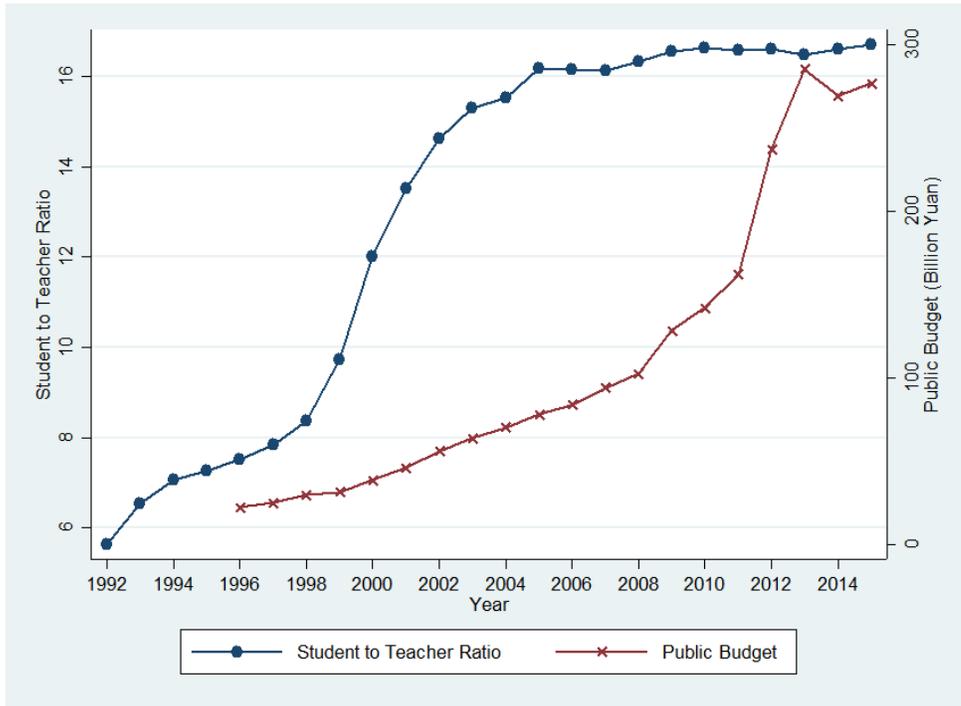
Note: The map plots the average annual increase rates of college admissions in each province of mainland China in 1998-2012. Provinces with blank fillings denote average increase rates less than 10%, provinces with horizontal lines denote average increase rates between 10%-13%, provinces with forward slashes denote average increase rates between 13%-15%, and provinces with vertical lines denote average increase rates greater than 15% (Source: Educational Statistics Yearbook of China)

Figure 2.7 Gross College Enrollment Ratio and Aged 18-22 Population in 1992-2015



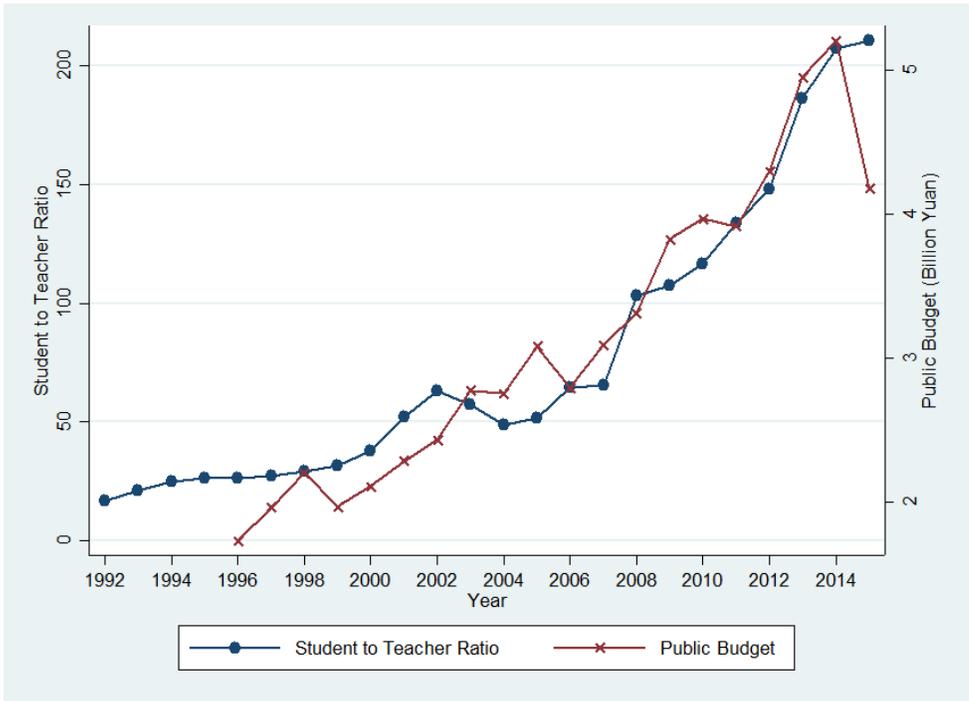
Note: This figure plots the gross college enrollment ratio and aged 18-22 population in 1992-2015. Due to lack of data, 18-22 population (million) is derived using the number of in-school higher education students, divided by gross enrollment rate for higher education (Source: Educational Statistics Yearbook of China).

Figure 2.10 National Public Budget and Ratio between Number of Students and Number of Teachers in Regular Colleges in 1992-2015



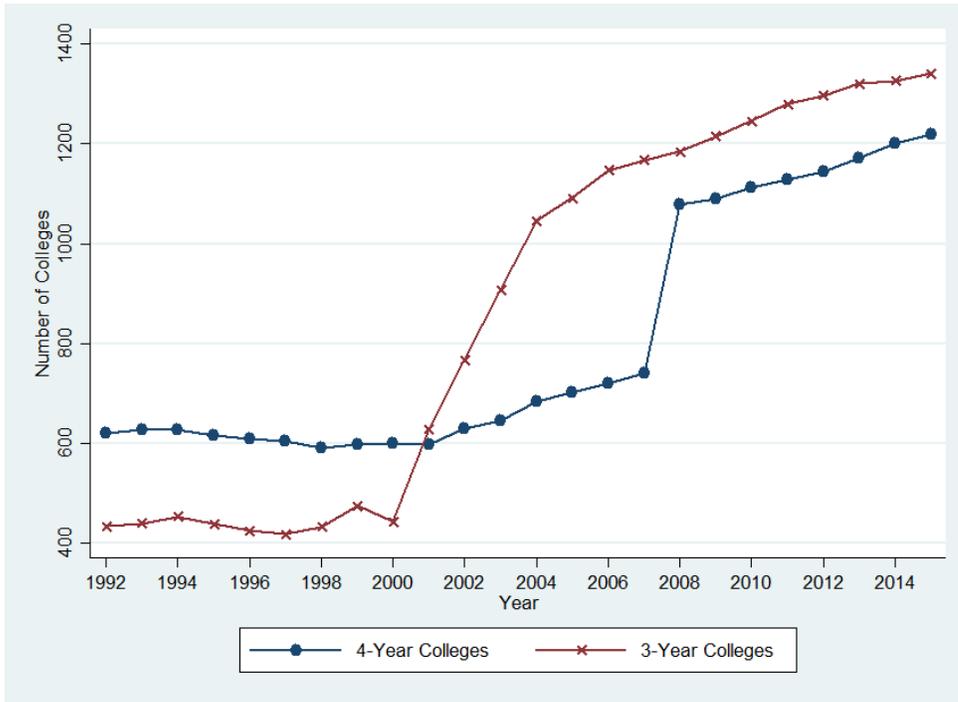
Note: This figure plots the national public budget for regular college education (dotted line, unit: billion Yuan) and ratio of number of students to number of teachers in regular colleges (cross line) (Source: Educational Statistics Yearbook of China).

Figure 2.13 National Public Budget and Ratio between Number of Students and Number of Teachers in Adult Colleges in 1992-2015



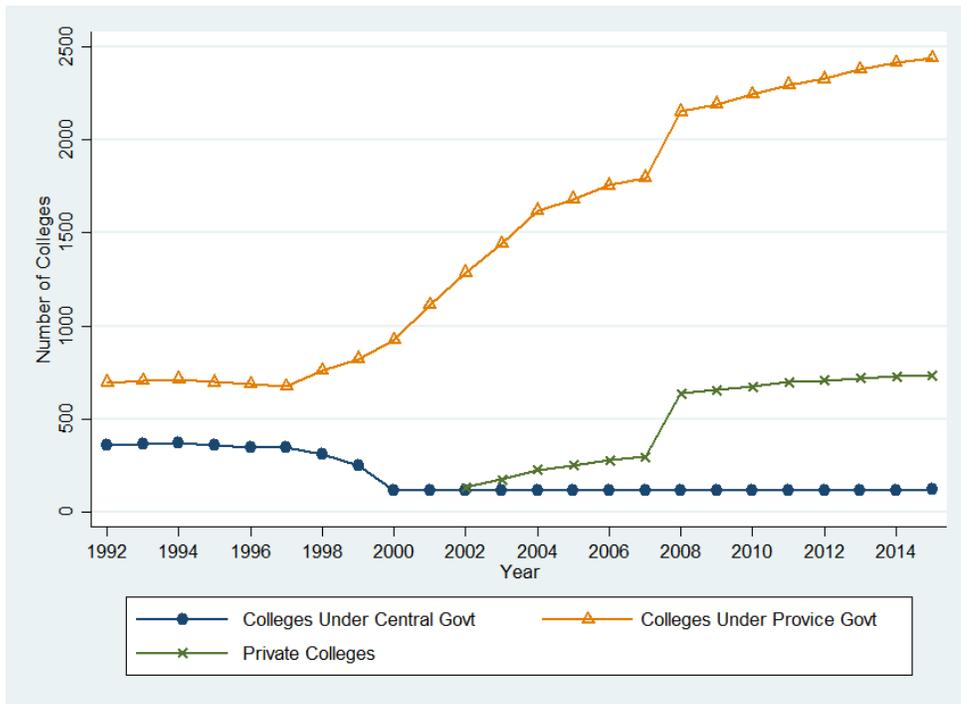
Note: This figure plots the national public budget for adult college education (dotted line, unit: billion Yuan) and ratio of number of students to number of teachers in regular colleges (cross line) (Source: Educational Statistics Yearbook of China).

Figure 2.16 Number of Colleges by Study in 1992-2015



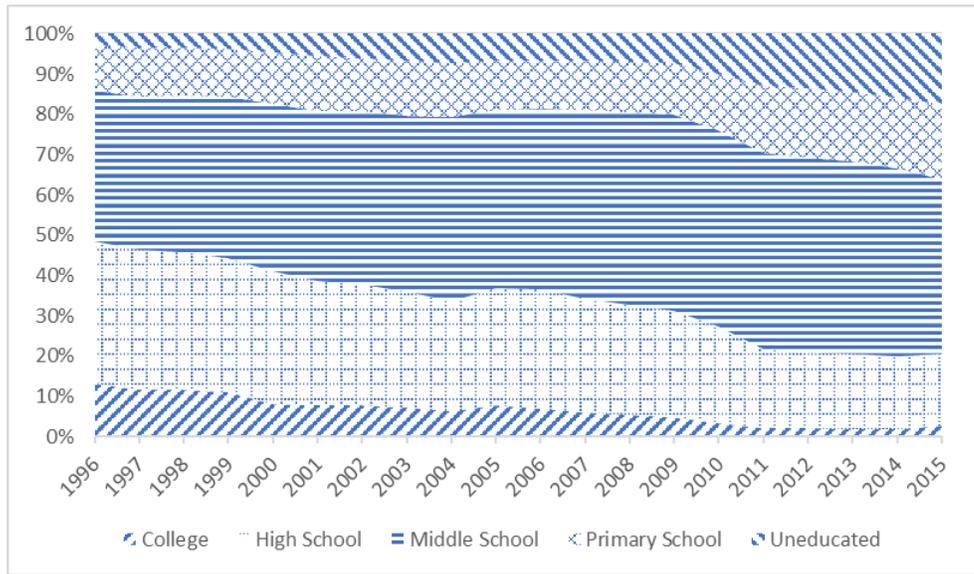
Notes: The graph shows the number of 4-year colleges (dotted line) and number of 2 or 3-year colleges (cross line) in China (Source: Educational Statistics Yearbook of China).

Figure 2.19 Number of Colleges by Control Agency in 1992-2015



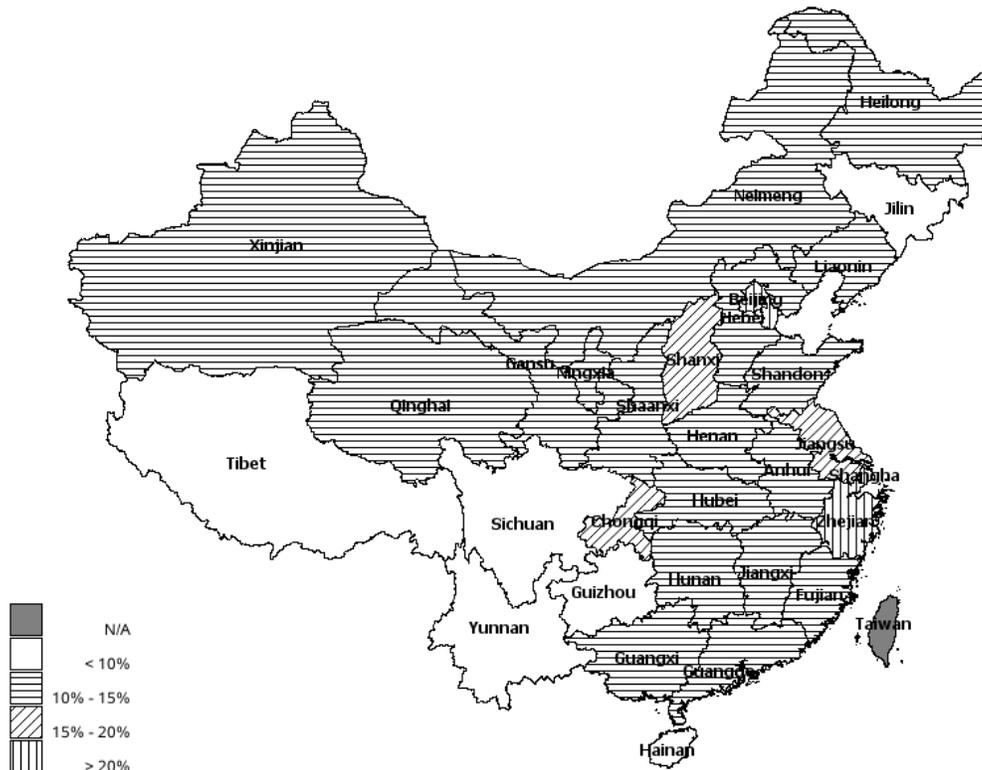
Notes: The graph shows the number of colleges under central government, under local agencies, and private colleges in China. Private colleges are under province government. Due to lack of statistics on private colleges in the education yearbooks, we plot the number of private colleges since 2002. The number of colleges under province in the graph also includes the number of private colleges. The sum of the number of colleges under central government and province government is equal to the sum of 4-year and 2 or 3-year colleges in Figure 2.6 (Source: Educational Statistics Yearbook of China).

Figure 2.22 Educational Distribution for Employed Workers in 1996-2015



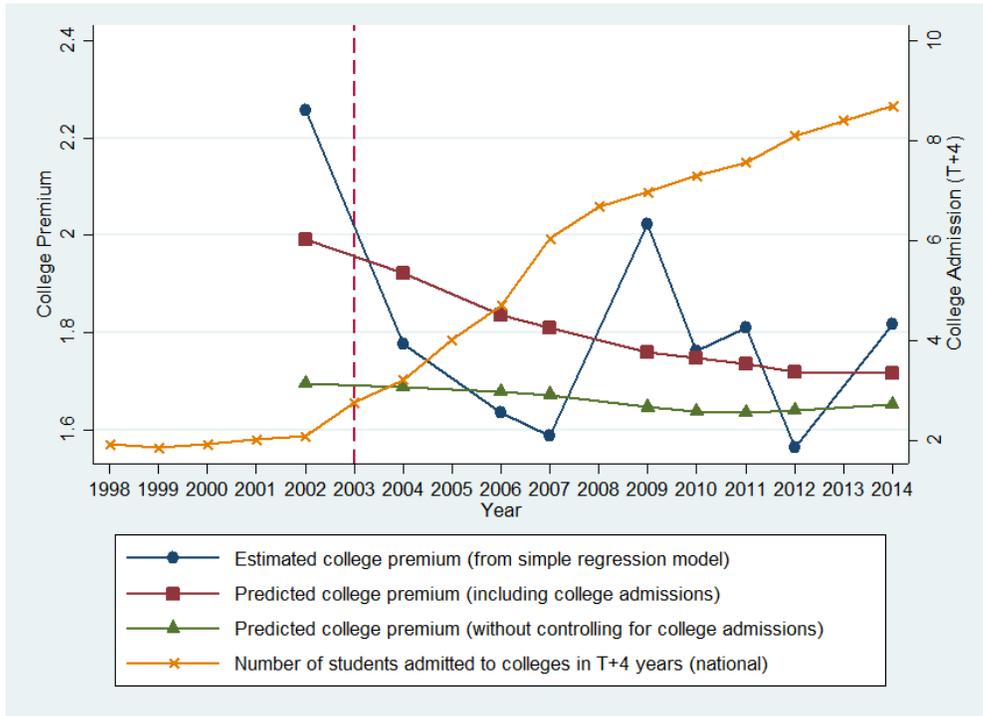
Note: The graph plots the educational distribution for employed workers in 1996-2015. The area with backslash denotes the proportion of employed workers with college degrees, the area with oblique square denotes the proportion of employed workers with high school degrees, the area with horizontal lines denotes the proportion of employed workers with middle school diplomas, the area with square denotes the proportion of employed workers with primary school diplomas, and the forward slash denotes the proportion of employed workers with less than primary school education (Source: China Labor Statistical Yearbook).

Figure 2.25 Change of Proportions in College Workers in China from 1998 to



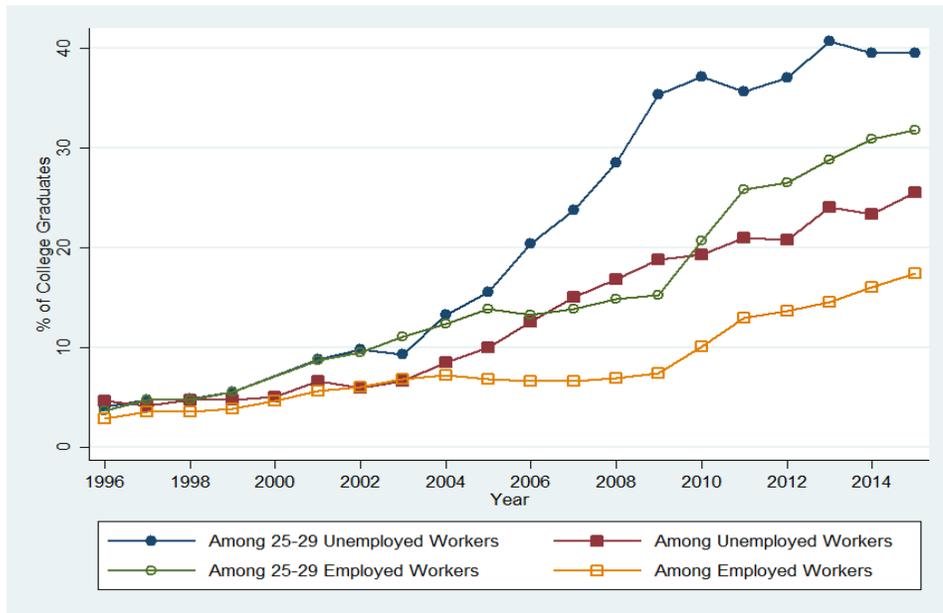
Note: This map plots the change of proportions in college workers in China from 1998 to 2012. Provinces with blank fillings denote average increase rates less than 10%, provinces with horizontal lines denote average increase rates between 10%-15%, provinces with forward slashes denote average increase rates between 15%-20%, and provinces with vertical lines denote average increase rates greater than 20% (Source: China Labor Statistical Yearbook).

Figure 2.28 Predicted College Premium and Number of Students Admitted to Colleges in 1998-2015



Note: The graph plots the predicted college premiums using different models, and the number of students admitted to colleges in 1998-2015. The predicted college premiums using simple model (regress income on college degree attainment) are shown in the dotted line, the predicted college premiums using the model without the number of college admissions but with controlled variables are shown in the triangle line, the predicted college premiums using the model with the number of college admissions but with controlled variables are shown in the square line, and the number of students admitted to colleges in $T+4$ year.

Figure 2.31 Proportions of College Graduates Among Employed and Unemployed Workers



Note: This figure plots the percentage of college graduates among urban unemployed workers (the solid-square line) and the percentage of college graduates among urban employed workers (the hollow-square line); as well as percentage of college graduates among 23-29-year-olds urban unemployed workers (the solid-circle line) and percentage of college graduates among urban 23-29-year-olds unemployed workers (the hollow-circle line) (Source: China Labor Statistical Yearbook).

2.7 Appendix

A1. Impact of the Skilled Labor Supply Shock on Labor Market Outcomes Using 2SLS Method

The changes in the number of skilled workers in a province can be explained by changes in both skilled labor supply and demand. When supply increases, more skilled workers enter labor market given the same wage levels. When demand increases, companies are willing to recruit more skilled workers at any given wage level. Therefore, using an instrument which measures external changes in supply of skilled labor can help to separate the changes in supply of skilled labor from the changes in skilled labor caused by changes in demand. In this section, we study the impact of a positive supply shock in skilled labor market on college premium and unemployment using 2SLS method.

In the first stage, we check whether the number of college admissions in each province and year is relevant to the change in skilled labor. We thus use the following model:

$$\Delta Col_{p,t} = \beta AD_{p,t-4} + \boldsymbol{\gamma}' \mathbf{Z}_{p,t-4} + \boldsymbol{\rho}' \mathbf{Z}_{p,t} + \delta_p + \delta_t + \varepsilon_{p,t} \quad (2.4)$$

$\Delta Col_{p,t}$ denotes the change in the number of skilled labor in year t and in province p , which is the difference between skilled labor in year t and $t-1$ ($\Delta Col_{p,t} = Col_{p,t} - Col_{p,t-1}$). $AD_{p,t-4}$ is our instrument for the supply shock, which is the number of students admitted to college in year $t-4$ in province p . $\mathbf{Z}_{p,t-4}$ and $\mathbf{Z}_{p,t}$ represents time-variant provincial demand-side factors, including the provincial log of GDP, the log of fixed capital investment, and the log of the imports and exports at year $t-4$ in province p , and at year t in province p , respectively. Including $\mathbf{Z}_{p,t-4}$ helps us to control the impact of previous demand change on both the college admissions and the current supply of skilled labor.

Including $Z_{p,t}$ helps to control the impact of the concurrent demand on the current supply of skilled labor. δ_p is the province fixed effects and δ_t is the year fixed effects. The provincial data for number of students admitted to college, GDP, import, export and fixed capital investment are obtained from the statistical yearbooks for the 31 provinces³⁰ in mainland China from 1998 to 2015.

Due to lacking the direct data on the number of skilled labor in each province ($\Delta Col_{p,t}$), we use other related data to infer the number of skilled labor in each province. The related statistics to infer the number of skilled labor in a province in each year are from the Chinese Statistical Yearbook and Chinese Educational Statistical Yearbook 1998-2015. The Chinese Statistical Yearbook surveyed about 1% of total population in each year except for 2000 and 2010. In 2000 and 2010, the government conducted census survey which covered all population except for servicemen.

The full procedure to obtain $\Delta Col_{p,t}$ from the yearbook is as follow: First, we calculate the percentage in college or with at least some college education among people older than 6 in each province and year from 1998-2015, $\%Col_{p,t}$, using the reported tables “Population by Sex, Educational Level and Region” from the yearbooks. Second, we obtain the percentage of people older than 6 in each province and year, $\%Age6_{p,t}$, using the reported tables “Household, Population and Sex Ratio by Region” from the yearbooks.

³⁰ We imputed the data for Chongqing and Sichuan provinces in 1996 because Chongqing was separated from Sichuan province and became as a municipality (a type of province) since 1997. To impute 1996’s values for Chongqing and Sichuan, we calculated the proportions of Sichuan and Chongqing in their summation values of 1997 for each of the variables in equation (2.1) and multiply these proportions with the 1996 Sichuan (which included Chongqing in that year) data.

Next, we multiply the resident population in each province and year³¹, $Pop_{p,t}$, with the percentage of population older than 6 and percentage of college graduates, $(Pop_{p,t} * \%Age6_{p,t} * \%Col_{p,t})$, to obtain the sum of the number of people who enrolled in colleges and people who received college education and already left the college (we will call these two groups of people “college attendees”).

Then, we subtract the number of people enrolling in colleges in each province and year, $Enr_{p,t}$ from the number of college attendees³² since it changed substantially since the college expansion. We then obtain the information of labor participation rate for college attendees older than 16 years old, $\%LPC_{p,t}$; as well as the number of people enrolling in colleges as a fraction of number of college attendees outside of labor force, $\%Enr_{p,t}$.³³ We assume the labor participation rate for college attendees, and the ratio between the college attendees outside of labor force other than enrolling in colleges versus the number of college attendees in the labor force are both stable across our observed year. Therefore, $\frac{\%Lpc}{(1-\%Lpc)(1-\%Enr)+\%Lpc}$ measures the number of college attendees in the labor force as a fraction of the number of college attendees who are not enrolled in colleges in year t . We then multiply the number of college attendees who are not enrolled in colleges with the

³¹ This data is obtained from the tables “Total Population and Birth Rate, Death Rate and Natural Growth Rate by Region” from the yearbooks.

³² These statistics is obtained from the Chinese Educational Statistical Yearbook.

³³ The labor participation rates for college attendees are obtained from the table “Employment Status Among 16 and above Population by Educational Level and Gender” in the 2010 census of each province; and the fractions of people enrolling in college in population outside of labor force are obtained from “The total number of people outside of labor force by reason, educational level, and gender” in 2010 census of each province.

ratio $\frac{\%Lpc}{(1-\%Lpc)(1-\%Enr)+\%Lpc}$ to obtain the estimates for number of skilled workers in province p in year t .

Finally, to capture the changes in the number of skilled labor in t of province p , we subtract the total number of skilled workers in the labor market in year $t-1$ from year t :

$$\Delta Col_{p,t} = \Delta[(Pop * \%Age6 * \%Col - Enr) * \frac{\%Lpc}{(1 - \%Lpc)(1 - \%Enr) + \%Lpc}]^{p,t} \quad (2.5)$$

In the second stage, we estimate the impact of the predicted changes in skilled labor due to the college expansion, $\Delta \widehat{Col}_{p,t}$, on the college premium and the probability of unemployment using equation (2.1) (replace $\widehat{AD}_{p,t-4}$ with $\Delta \widehat{Col}_{p,t}$).

The results for the first stage are reported in APTable 2.7, and we also report the estimated results using different set of control variables. Column (1) shows the estimated coefficient for the instrument without any control variable. Column (2) adds province fixed effects and year fixed effects in the model. Column (3) further control demand variables (GDP, import values, export values, fixed capital investment values) of the current year t and province p , as well as demand statuses in $t-4$ year at province p . We use the predicted results from the last specification in APTable 2.7 to do the second stage estimation.

We estimate the relationship between the college admissions in $t-4$ year and the change of skilled labor in the market in year t . As can be seen from the table, an increase of one student in college admission increases the number of skilled workers by about 0.99, controlling for the previous and current changes in several the demand factors. The

estimated coefficient suggests the local college admissions 4-year prior can explain a substantial portion of the variation in skilled labor. Compared with results in columns (1) and (2), we can see that adding province fixed effects, year fixed effects, and demand side variables does not significantly affect the estimated relationship between college admission and actual changes in the number of skilled labor. Therefore, our instrument is external to demand changes, fixed provincial changes and fixed year changes in terms of affecting the number of skilled labor.

The results of the second stage regression on the college premium and the probability of unemployment can be found in APTable 2.8 and APTable 2.9. Columns (1) – (3) and columns (5) – (7) in APTable 2.8 report the regression results using the predicted change in the supply of skilled labor from the first stage, $\Delta\widehat{Col}_{p,t}$ and its interactions with different educational dummies, as the independent variables of interested. Columns (4) and (8) report the regression results using raw number of $\Delta Col_{p,t}$ and its interactions with different education dummies as the main independent variables.

In APTable 2.8, we observe that the instrumented changes in skilled labor due to the college expansion had a negative impact on the relative income for young college workers. A 0.1 million increase in skilled workers from the college expansion decreased the relative log income for 4-year college graduates by about 3 log percentage points. This indicates that the change in number of skilled workers due to the college expansion decreased the college premium for young skilled labor. It also suggests the price elasticity of demand for

skilled labor is about -32^{34} , which suggests a highly elastic demand for skilled labor in Chinese market. Compared with the estimated impact of the college expansion in Table 2.5, the estimated impact of supply shock on college premium using the 2SLS is smaller. The results in column (4) suggests that the impact of the raw change in number of skilled workers on college premium is close to zero. Since the raw change in skilled labor may also be affected by changes in the demand of skilled labor, such as greater usage in skilled bias technology may cause the college premium to increase, the comprehensive impact of change in skilled workers on college premium is thus ambiguous.

The increase in skilled workers due to supply shock has no negative impact on college premium for older workers, as can be seen in columns (5) – (7), which suggests that the effect of having greater number of skilled workers in the market mainly affects young people rather than older people. In column (8), we find that the income premiums for both older 4-year and 2- or 3-year college graduates increase as the number of skilled workers increases. This suggests an increase in demand for experienced skilled labor, which is consistent with the skill-biased technological changes theory.

In APTable 2.9, we estimate the impact of changes in skilled labor due to the college expansion on the likelihood of unemployment. Similar to APTable 2.8, columns (1) – (3) and columns (5) – (7) show the estimated results using instrumented variable, and columns (4) and (8) show the results using raw changes in skilled labor as our independent variable of interest. The results suggest that as the skilled labor increases due to the college

³⁴ This number is calculated using the percentage change in quantity, $(1/1.05)*100\%$, where the 1.05 is the average changes in quantity of skilled labor; divided by the percentage change in price, -3% , which is the change in college premium from the regression results for 0.1 million increase in college admission.

expansion, the likelihood of unemployment for both young and older college graduates increases relative to high school graduates. A 0.1 million increase in local skilled labor caused by the college expansion increases the probability of unemployment for 4-year new college graduates by 1.1 percentage points, and it increased the probability for old graduates by 0.9 percentage points, relative to people with high school degrees. It increased the relative probability of unemployment for 2- or 3-year new college graduates by about 2 percentage points and it increased the relative probability by 1.2 percentage points for older college graduates. Similar to the results in APTable 2.8, the negative impacts on both young and old college graduates are smaller than the reduced form estimates.

The comparison between the 2SLS estimates and the reduced form estimates on both relative income and probability of unemployment suggests that the overall negative impact of the college expansion on the earning premium and employment for young skilled labor is greater than the estimate using college expansion as an instrument for the number of skilled workers. In other words, the increase in college admission negatively affected young skilled workers' income and probability of employment through increasing the number of skilled workers in the market, it may have negatively affected their labor outcomes through other channels. For example, a decline in college quality may decrease the demand for skilled labor. As shown in Section 3, during the college expansion period, the increase in number of students admitted to colleges grew faster than the increase in college teachers, which may result in a decrease in the quality of college education. If college graduates are on average not as qualified as before, which may decrease their

marginal product, then the employer would not provide wages as high as before and may also decrease the demand for skilled labor.

The estimated increase in relative probability of unemployment for college graduates using the raw change in skilled labor are smaller than the 2SLS estimates. Since an increase in number of skilled workers is a mixture results of the changes in both supply and demand, the estimated impacts using the simple change is therefore ambiguous and cannot be identified as either the impact of supply increase or impact of demand increase.

APTable 2.1 Regression Results for College Premium within Industry and within Occupation

	Age in 23-29				Age in 30-65			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Col	0.552*** (0.110)	0.581*** (0.114)	0.501*** (0.114)	0.514*** (0.120)	0.603*** (0.052)	0.618*** (0.057)	0.539*** (0.056)	0.540*** (0.057)
Lcol	0.171** (0.074)	0.204*** (0.077)	0.125 (0.084)	0.144* (0.086)	0.286*** (0.032)	0.294*** (0.034)	0.224*** (0.034)	0.224*** (0.034)
Lhs	-0.162* (0.095)	-0.157* (0.091)	-0.162* (0.093)	-0.161* (0.093)	-0.308*** (0.033)	-0.302*** (0.034)	-0.239*** (0.033)	-0.237*** (0.033)
\widetilde{AD}_{t-4}	-0.044*** (0.012)	0.012 (0.075)	-0.015 (0.075)	-0.011 (0.076)	0.042 (0.031)	0.036 (0.031)	0.055* (0.031)	0.049 (0.030)
Col * \widetilde{AD}_{t-4}	-0.203*** (0.060)	-0.200*** (0.057)	-0.196*** (0.056)	-0.198*** (0.058)	-0.015 (0.029)	-0.019 (0.029)	-0.029 (0.029)	-0.027 (0.029)
Lcol * \widetilde{AD}_{t-4}	-0.035 (0.048)	-0.043 (0.046)	-0.028 (0.046)	-0.035 (0.046)	-0.003 (0.020)	0.003 (0.020)	0.001 (0.020)	0.004 (0.020)
Lhs * \widetilde{AD}_{t-4}	-0.016 (0.052)	-0.030 (0.049)	-0.012 (0.050)	-0.017 (0.050)	0.008 (0.018)	0.008 (0.018)	0.002 (0.018)	0.002 (0.018)
X_i	Yes							
δ_p	Yes							
\overline{TL}	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
$Z_{p,t}$	Yes	Yes	Yes	Yes	No	No	Yes	Yes
δ_t	Yes	Yes	Yes	Yes	No	No	No	Yes
δ_i	No	Yes	No	Yes	No	Yes	No	Yes
δ_o	No	No	Yes	Yes	No	No	Yes	Yes
N	3,735	3,735	3,735	3,735	16,852	16,852	16,852	16,852
R ²	0.2739	0.2832	0.3002	0.3087	0.3424	0.3497	0.3734	0.3785

Note: This table reports the regression results for equation (2.1), controlling for industry fixed effects and occupation fixed effects, using CGSS 2003, 2006, 2010, 2011, 2012, 2013. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Columns (1)-(4) list estimates for young workers (23-29-year-olds), and columns (5)-(8) list estimates for older workers (30-65-year-olds). We control industry fixed effects in columns (2) and (6), control occupation fixed effects in columns (3) and (7), and control both industry and occupation effects in columns (4) and (8). The other control variables are the same as described in Table 2.5. Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

APTable 2.2 Regression Results Based on Equations (2.2) and (2.3)

Panel A: Results on Equation (2.2)				
$GDP_{p,T-1}$	0.310***			
δ_p	YES			
N	588			
R^2	0.8002			
Panel B: Results on Equation (2.3)				
	Age in 23-29		Age in 30-65	
	Income (1)	Unemployment (2)	Income (3)	Unemployment (4)
Col	0.387*** (0.051)	-0.074*** (0.011)	0.649*** (0.028)	-0.087*** (0.009)
Lcol	0.168*** (0.049)	-0.044*** (0.011)	0.291*** (0.022)	-0.074*** (0.005)
Lhs	-0.141*** (0.078)	0.073*** (0.015)	-0.262*** (0.024)	0.066*** (0.006)
ε_{t-4}	0.041 (0.108)	-0.004 (0.026)	0.096* (0.050)	-0.012 (0.014)
Col * ε_{t-4}	-0.116** (0.046)	0.028* (0.017)	0.066*** (0.025)	0.024*** (0.009)
Lcol * ε_{t-4}	0.035 (0.049)	-0.010 (0.018)	-0.017 (0.023)	0.008 (0.009)
Lhs * ε_{t-4}	-0.055 (0.073)	-0.134*** (0.031)	-0.076*** (0.022)	-0.093*** (0.012)
X_i	Yes	Yes	Yes	Yes
δ_p	Yes	Yes	Yes	Yes
$\tilde{T}L$	Yes	Yes	Yes	Yes
$Z_{p,t}$	Yes	Yes	Yes	Yes
δ_t	Yes	Yes	Yes	Yes
N	3,572	5,334	16,852	24,738
R^2	0.4968	0.0879	0.4720	0.1540

Note: This table reports the regression results for equations (2.2) and (2.3). Panel A lists the regression results for equation (2.2), where the dependent variable is the demeaned number of college admissions in province p in year T ($T \in 1997 - 2012$). Independent variable is the GDP in province p in year $T-1$. Panel B lists the regression results for equation (2.3). The dependent variable is the log of annual income for full-time urban employed respondents (columns (1) and (3)) or dummy for unemployment for urban respondents (columns (2) and (4)). $\varepsilon_{p,t-4}$ is the residual from equation (2.2) at year $t-4$, where t is the year when individual i took the survey. The other variables are the same as described in Table 2.5. Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

APTable 2.3 Regression Results Based on Equation (2.1) Using the Ratio between Provincial College Admission and Population

	Age in 23-29		Age in 30-65	
	Income (1)	Unemployment (2)	Income (3)	Unemployment (4)
Col	0.384*** (0.042)	-0.054*** (0.014)	0.603*** (0.056)	-0.072*** (0.008)
Lcol	0.147*** (0.038)	-0.065*** (0.015)	0.290*** (0.024)	-0.080*** (0.008)
Lhs	-0.124** (0.050)	0.066*** (0.017)	-0.333*** (0.025)	0.062*** (0.008)
$\frac{AD}{POP_{t-4}}$	1.263*** (0.239)	-0.280*** (0.045)	1.031*** (0.149)	-0.453*** (0.065)
Col * $\frac{AD}{POP_{t-4}}$	-0.138* (0.059)	0.054*** (0.016)	0.028 (0.067)	0.019 (0.015)
Lcol * $\frac{AD}{POP_{t-4}}$	-0.076 (0.074)	0.057** (0.026)	0.071 (0.052)	0.018 (0.017)
Lhs * $\frac{AD}{POP_{t-4}}$	-0.032 (0.064)	-0.097** (0.040)	0.104*** (0.032)	-0.062* (0.036)
X_i	Yes	Yes	Yes	Yes
δ_p	Yes	Yes	Yes	Yes
$Z_{p,t}$	Yes	Yes	Yes	Yes
δ_t	Yes	Yes	Yes	Yes
N	3,572	5,334	16,852	24,738
R ²	0.3360	0.0952	0.4091	0.1352

Note: This table reports the regression results on equation (2.1) using the ratio between provincial college admission and population as the measure of the degree of the college expansion ($\frac{AD}{POP_{t-4}}$). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. All of other variables are the same as the variables shown in Table 2.5 and Table 2.6. Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

APTable 2.4 Placebo Test on the Impact of the College Expansion on Other Macro-Level Variables

	Population (1)	GDP (2)	Unemployment Rate (3)
AD_{t-4}	-0.017 (0.034)	-0.021 (0.017)	-0.003 (0.036)
y_{t-1}	Yes	Yes	Yes
y_{t-2}	Yes	Yes	Yes
y_{t-3}	Yes	Yes	Yes
y_{t-4}	Yes	Yes	Yes
N	620	620	620
R ²	0.9980	0.9771	0.5483

Note: This table reports the estimated results of placebo tests on the impact of the college expansion on population, GDP, and unemployment rate. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variables include provincial population, GDP, unemployment in year t , the variable of interest is the number of provincial college admission in year $t-4$. Control variables include the first, second, third, and fourth lags of the dependent variable.

APTable 2. 5 Marginal Effect Estimates Using Logit Model on Unemployment

	Age in 23-29			Age in 30-65		
	(1)	(2)	(3)	(4)	(5)	(6)
Col	-0.074*** (0.016)	-0.069*** (0.017)	-0.065*** (0.016)	-0.204*** (0.021)	-0.175*** (0.020)	-0.169*** (0.020)
Lcol	-0.080*** (0.013)	-0.077*** (0.013)	-0.075*** (0.013)	-0.126*** (0.014)	-0.118*** (0.011)	-0.118*** (0.011)
Lhs	0.046*** (0.011)	0.046*** (0.011)	0.045*** (0.011)	0.048*** (0.005)	0.045*** (0.005)	0.045*** (0.005)
\widetilde{AD}_{t-4}	-0.025* (0.013)	-0.011 (0.015)	-0.017 (0.019)	-0.073*** (0.010)	0.003 (0.015)	0.028** (0.011)
Col * \widetilde{AD}_{t-4}	0.042*** (0.012)	0.042*** (0.013)	0.041*** (0.013)	0.032*** (0.011)	0.010 (0.013)	0.008 (0.013)
Lcol * \widetilde{AD}_{t-4}	0.046*** (0.013)	0.046*** (0.012)	0.044*** (0.014)	0.021* (0.011)	0.010 (0.013)	0.010 (0.010)
Lhs * \widetilde{AD}_{t-4}	-0.029** (0.014)	-0.029** (0.013)	-0.027** (0.012)	-0.016*** (0.005)	-0.019*** (0.005)	-0.019*** (0.005)
X_i	Yes	Yes	Yes	Yes	Yes	Yes
δ_p	Yes	Yes	Yes	Yes	Yes	Yes
\widetilde{TL}	No	Yes	Yes	No	Yes	Yes
$Z_{p,t}$	No	Yes	Yes	No	Yes	Yes
δ_t	No	No	Yes	No	No	Yes
N	5,334	5,334	5,334	24,738	24,738	24,738
R ²	0.1114	0.1153	0.1282	0.1507	0.1697	0.1822

Note: This table reports the marginal effect obtained from Logit regression estimates for equation (2.1). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. All variables listed in the table have the same definitions as shown in Table 2.6. Columns (1)-(3) list estimates for young workers, and columns (4)-(6) list estimates for older workers. We control the province fixed effects in columns (1) and (4), adding provincial economic variables in columns (2) and (5), and including year fixed effects in columns (3) and (6). Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

APTable 2.6 Estimates of Returns to Task Scores within Education

In(annual income)	(1)	(2)	(3)	(4)
Cognitive Analytical	1.325*** (0.110)	--	--	--
Cognitive Communicative	--	0.320*** (0.022)	--	--
Manual Precision	--	--	-0.252*** (0.024)	--
Manual Others	--	--	--	-0.759*** (0.069)
Col	0.504*** (0.024)	0.458*** (0.024)	0.529*** (0.023)	0.532*** (0.023)
Lcol	0.204*** (0.020)	0.169*** (0.021)	0.217*** (0.020)	0.218*** (0.020)
Lhs	-0.248*** (0.019)	-0.225*** (0.019)	-0.262*** (0.019)	-0.263*** (0.019)
Gender	0.263*** (0.015)	0.278*** (0.015)	0.274*** (0.015)	0.276*** (0.015)
Age	0.026*** (0.007)	0.027*** (0.007)	0.026*** (0.007)	0.027*** (0.007)
Age * Age	-0.0004*** (0.0001)	-0.0004*** (0.0001)	-0.0004*** (0.0001)	-0.0004*** (0.0001)
δ_p	Yes	Yes	Yes	Yes
δ_t	Yes	Yes	Yes	Yes
N	20,587	20,587	20,587	20,587
R ²	0.2678	0.2715	0.2662	0.2661

Note: This table reports the estimates for returns to task scores within education in current jobs for people in 25-65 years old. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is log of annual income one year before the survey year for respondents with full-time employment. Our variable of interest is the score for each type of task for the respondent. The data is from CGSS 2003-2015, excluding 2005. We control province fixed effects and year fixed effects, as well as education dummies, gender, age and age squared in our estimation. Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

APTable 2.7 Estimated Results of the First Stage in the 2SLS Regression

$\Delta Col_{p,t}$	(1)	(2)	(3)
$AD_{p,t-4}$	0.972*** (0.148)	0.980*** (0.376)	0.994** (0.421)
δ_p	No	Yes	Yes
δ_t	No	Yes	Yes
$Z_{p,t-4}$	No	No	Yes
$Z_{p,t}$	No	No	Yes
F-score	44.03***	23.93***	20.11***
N	527	527	527
R ²	0.0756	0.2344	0.2384

Note: This table reports the results of the first stage of the 2SLS analysis described in equation (2.4), as well as other two specifications with less control variables. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is the change in skilled labor in year t ($t \in [1999, 2015]$) and in province p . $AD_{p,t-4}$ is the number of students admitted to college in year $t-4$ in province p . $Z_{p,t-4}$ and $Z_{p,t}$ represents time-variant provincial demand-side factors, including the provincial log of GDP, the log of fixed capital investment, and the log of the imports and exports at year $t-4$ in province p , and at year t in province p , respectively. δ_p is the province fixed effects and δ_t is the year fixed effects. Column (1) shows the estimated coefficient for the instrument without any control variable. Column (2) shows the estimated coefficient for the instrument with province fixed effects and year fixed effects controlled in the model. Column (3) shows the estimated coefficient for the instrument with controls in column (2) and with demand variables in the current year t and province p , as well as demand status in $t-4$ year.

APTable 2.8 Estimated Results of the Second Stage in the 2SLS Regression on Income

	Age in 23-29				Age in 30-65			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Col	0.445*** (0.034)	0.382*** (0.035)	0.376*** (0.033)	0.361*** (0.029)	0.624*** (0.047)	0.583*** (0.043)	0.579*** (0.042)	0.598*** (0.045)
Lcol	0.188*** (0.033)	0.156*** (0.031)	0.157*** (0.031)	0.143*** (0.030)	0.340*** (0.029)	0.332*** (0.026)	0.329*** (0.025)	0.334*** (0.025)
Lhs	-0.070 (0.051)	-0.145*** (0.048)	-0.144*** (0.049)	-0.148*** (0.049)	-0.265*** (0.021)	-0.288*** (0.022)	-0.289*** (0.021)	-0.291*** (0.019)
$\widehat{\Delta Col}_t$	-0.028** (0.011)	-0.026** (0.011)	-0.029 (0.029)	-0.001 (0.004)	-0.032*** (0.009)	-0.020** (0.009)	0.005 (0.040)	-0.007*** (0.002)
Col * $\widehat{\Delta Col}_t$	-0.037*** (0.012)	-0.032** (0.013)	-0.028** (0.012)	-0.001 (0.006)	0.012 (0.014)	0.010 (0.012)	0.009 (0.012)	0.009*** (0.003)
Lcol * $\widehat{\Delta Col}_t$	-0.030* (0.015)	-0.026* (0.014)	-0.024* (0.014)	0.003 (0.005)	0.013 (0.009)	0.011 (0.009)	0.011 (0.009)	0.003 (0.002)
Lhs * $\widehat{\Delta Col}_t$	-0.012 (0.020)	-0.010 (0.018)	-0.012 (0.018)	-0.003 (0.005)	0.001 (0.009)	0.001 (0.009)	0.001 (0.001)	0.003 (0.002)
X_i	Yes							
δ_p	Yes							
\overline{TL}	No	Yes						
$Z_{p,t}$	No	Yes	Yes	Yes	No	Yes	Yes	Yes
δ_t	No	No	Yes	Yes	No	No	Yes	Yes
N	3,735	3,735	3,735	3,735	16,852	16,852	16,852	16,852
R ²	0.3245	0.3761	0.3824	0.3821	0.3794	0.4061	0.4101	0.4105

Note: This table reports the second stage results on annual incomes for the 2SLS regression using equation (2.1), replacing $\widehat{AD}_{p,t-4}$ with $\widehat{\Delta Col}_{p,t}$. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is the log of annual income for full-time urban employed respondents. Our variable of interest is the predicted change in skilled workers caused by the supply shock in year t of province p , and its interaction terms with education dummies. Columns (1)-(4) list estimates for young workers, and columns (5)-(8) list estimates for older workers. We control the age, age squared, gender, province fixed effects, effect of changes in total employed workers in columns (1) and (5), adding provincial economic and demographic variables in columns (2) and (6), and including year fixed effects in columns (3) and (7). We also report the estimated results using raw number of changes in college graduate workers in columns (4) and (8). Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

APTable 2.9 Estimated Results of the Second Stage in the 2SLS Regression on Unemployment

	Age in 23-29				Age in 30-65			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Col</i>	-0.051*** (0.010)	-0.035*** (0.011)	-0.036*** (0.009)	-0.029*** (0.009)	-0.097*** (0.008)	-0.072*** (0.007)	-0.069*** (0.006)	-0.059*** (0.005)
<i>Lcol</i>	-0.060*** (0.010)	-0.054*** (0.010)	-0.056*** (0.010)	-0.047*** (0.011)	-0.086*** (0.007)	-0.083*** (0.006)	-0.087*** (0.006)	-0.078*** (0.006)
<i>Lhs</i>	0.054*** (0.017)	0.060*** (0.017)	0.055*** (0.017)	0.052*** (0.016)	0.065*** (0.008)	0.066*** (0.007)	0.063*** (0.006)	0.057*** (0.005)
$\widehat{\Delta Col}_t$	-0.007 (0.007)	-0.001 (0.007)	-0.045 (0.033)	-0.001 (0.002)	-0.005 (0.003)	0.002 (0.004)	-0.013 (0.025)	0.003 (0.002)
<i>Col</i> * $\widehat{\Delta Col}_t$	0.012* (0.007)	0.012* (0.007)	0.011* (0.007)	0.001 (0.002)	0.018*** (0.003)	0.009*** (0.002)	0.009*** (0.003)	0.001 (0.001)
<i>Lcol</i> * $\widehat{\Delta Col}_t$	0.018* (0.009)	0.016* (0.009)	0.019** (0.008)	0.005* (0.003)	0.012*** (0.004)	0.011*** (0.004)	0.013*** (0.004)	0.001 (0.001)
<i>Lhs</i> * $\widehat{\Delta Col}_t$	-0.020** (0.009)	-0.021** (0.008)	-0.017* (0.008)	-0.007 (0.004)	-0.017*** (0.003)	-0.017*** (0.003)	-0.017*** (0.003)	-0.004** (0.001)
X_i	Yes							
δ_p	Yes							
\widetilde{L}	No	Yes						
$Z_{p,t}$	No	Yes	Yes	Yes	No	Yes	Yes	Yes
δ_t	No	No	Yes	Yes	No	No	Yes	Yes
N	5,334	5,334	5,334	5,334	24,738	24,738	24,738	24,738
R ²	0.0545	0.0712	0.0874	0.0828	0.0888	0.1288	0.1494	0.1474

Note: This table reports the second stage results on the likelihood of unemployment using equation (2.1), replacing $\widehat{AD}_{p,t-4}$ with $\widehat{\Delta Col}_{p,t}$. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable is the log of annual income for full-time urban employed respondents. Our variable of interest is the predicted change in skilled workers caused by the supply shock in year t of province p , and its interaction terms with education dummies. The regression excludes people with more than 4-year college education. Columns (1)-(4) list estimates for young workers, and columns (5)-(8) list estimates for older workers. We control the age, age squared, gender, province fixed effects, effect of changes in total employed workers in columns (1) and (5), adding provincial economic and demographic variables in columns (2) and (6), and including year fixed effects in columns (3) and (7). We also report the estimated results using raw number of changes in college graduate workers in columns (4) and (8). Standard errors reported in this table are clustered at province level. The t-statistics are estimated using wild cluster bootstrap method (boottest).

2.8 Reference

Aamodt, Per Olaf, and Clara Åse Arnesen. "The Relationship Between Expansion in Higher Education and the Labour Market in Norway." *European Journal of Education* 30.1 (1995): 65-76.

Acemoglu, Daron. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *The Quarterly Journal of Economics* 113.4 (1998): 1055-1089.

Acemoglu, Daron. "Equilibrium Bias of Technology." *Econometrica* 75.5 (2007): 1371-1409.

Angrist, Joshua D., and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2008.

Department of Development & Planning Ministry of Education, P.R.C. (Year of 1992-2015 Edition). *Educational Statistics Yearbook of China* [Electronic Version]. Retrieved from China Labor Employment and Socioeconomic Development Statistical Database.

Department of Population and Employment Statistics National Bureau of Statistics, Department of Planning and Finance, and Ministry of Labor and Social Security, P.R.C. (Year of 1997-2015 Edition). *China Labor Statistical Yearbook* [Electronic Version]. Retrieved from China Labor Employment and Socioeconomic Development Statistical Database.

Dong, Wei. "The Difficulty for College Students to Find a Job Does Not Change." *China Youth Daily*, 02 December, 2012, http://zqb.cyol.com/html/2012-12/20/nw.D110000zgqnb_20121220_6-01.htm. In Chinese, accessed August 01, 2020.

Freeman, Richard B. "Overinvestment in College Training?" *Journal of Human Resources* (1975): 287-311.

Freeman, Richard Barry. "The Decline in the Economic Rewards to College Education." *The Review of Economics and Statistics* (1977): 18-29.

Goldin, Claudia, and Lawrence F. Katz. "Long-run Changes in the US Wage Structure: Narrowing, Widening, Polarizing." No. w13568. National Bureau of Economic Research, 2007.

He, Yu, and Yinhua Mai. "Higher Education Expansion in China and the 'Ant Tribe' Problem." *Higher Education Policy* 28.3 (2015): 333-352.

Katz, Lawrence F., and Kevin M. Murphy. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *The Quarterly Journal of Economics* 107.1 (1992): 35-78.

Li, Shi, John Whalley, and Chunbing Xing. "China's Higher Education Expansion and Unemployment of College Graduates." *China Economic Review* 30 (2014): 567-582.

Office of Population Census at the State Council of P.R.C. and Department of Population and Employment Statistics National Bureau of Statistics of P.R.C. (Year of 2000,2010 Edition). *Population Census of the P.R.C.* [Electronic Version]. Retrieved from China Statistics Bureau Database.

Ministry of Education of the People's Republic of China. "Measures for the Establishment and Administration of Independent Colleges". *MOE website*, 01 April, 2008, http://www.gov.cn/flfg/2008-03/07/content_912242.htm. In Chinese, accessed August 01, 2020.

Ministry of Education of the People's Republic of China. "Several Opinions of the MOE on Improving the Overall Quality of Higher Education." *MOE website*, 06 March, 2012, http://www.moe.gov.cn/srcsite/A08/s7056/201203/t20120316_146673.html. In Chinese, accessed August 01, 2020.

Pan, Zhixian. "Easy to Obtain Admission to Colleges but Hard to Find Jobs Reduces the Motivation for Young People to Go to Colleges." *People Website*, 14 March, 2005, <http://edu.people.com.cn/GB/1055/3240291.html>. In Chinese, accessed August 01, 2020.

Roodman, David, et al. "Fast and Wild: Bootstrap Inference in Stata Using boottest." *The Stata Journal* 19.1 (2019): 4-60.

Tang, Min, and Zuo, Xiaolei. "A Suggestion for Doubling the College Enrollment." *Economics Information*, 19 February.1999, <http://finance.sina.com.cn/review/20041023/15201102716.shtml>. In Chinese, accessed July 18, 2020.

Welch, Finis. "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust." *Journal of Political Economy* 87.5, Part 2 (1979): S65-S97.

Xu, Fanzeng. "Jiantao Ren: College Graduates Employment is A National Tragedy." *Southern People Weekly*, 04 June. 2007,

<http://business.sohu.com/20070604/n250386577.shtml>. In Chinese, accessed August 01, 2020.

Yao, Junyan. “College Expansion in China and Its Effect on Economic Outcomes”, Dissertation of University of North Carolina at Chapel Hill.

Yue Chen. “The Bankruptcy of Chinese Higher Education Economics.” *Directors & Boards*, 07 February, 2006,

<http://finance.sina.com.cn/magage/sxygl/20060207/16152323091.shtml>. In Chinese, accessed August 01, 2020.

Zhang, Chunxu. “Is Chines Universities Producing an Army of Unemployed Workers?” *Tencent Review*, 05 June, 2013, <https://view.news.qq.com/zt2013/uniworth/index.htm>. In Chinese, accessed August 01, 2020.

Zhang, Shuo. “MOE Set the Agenda for the 2017 Higher Education Admission.” *People.cn*, 11 May, 2017, <http://edu.people.com.cn/n1/2017/0511/c1053-29267405.html>. In Chinese, accessed August 01, 2020.

Chapter 3

The Impact of College Expansion on Internal Migration in China

3.1 Introduction

The college expansion in China (1999) has provided increasing opportunities for people to attend college and greatly upgraded China's labor force. In conjunction with increasing migration due to China joining the WTO, this policy has had an important effect on migration patterns in China. More college enrollment opportunities in both local province and provinces outside the local province change the migration behavior for people who plan to attend college. After graduating from colleges, competition from new college graduates entering the local job market then again affects their migration decisions. The college expansion also changes the migration behavior for those who do not attend college.

This chapter studies the impact of the college expansion on inter-provincial migration in China. It extends the previous migration literature on college-bound students (Tuckman, 1970; Fenske, 1972; Steahr and Lowe, 1975; Mixon and Hsing, 1994a; Mak and Moncur, 2003; Crook and Boyle, 2011; Faggian and Franklin, 2014; Liu et al., 2017) by studying the effect of college enrollment opportunities on migration for both college-bound students and who did not attend college. Also, we extend the literature on the migration of college graduates (Yousefi and Rives, 1987; Kodrzycki, 2001; Tornatzky et al., 2001; Gottlieb and

Joseph, 2006; Ishitani, 2010) by studying the competition effect from college expansion, as well as the effect of increasing the skilled labor supply on non-college graduates. In contrast to the previous literature on the causal effect of education on migration (Malamud and Wozniak, 2012; Machin et al., 2012; McHenry, 2013), this chapter argues that it is necessary to consider both effects on migration through the direct impacts of educational policies on the labor market (which influences both college graduates and others), and the impacts on those who attend college due to those policies. Papers which use college expansion as an instrument for individual educational attainment may be biased since the college expansion has effects through both channels. The chapter is structured as follows: Section 3.2 introduces the related migration literature; Section 3.3 describes the basic statistics of migration patterns, and the college expansion in China; Section 3.4 illustrates the model we use in this chapter; Section 3.5 shows the regression results and analyses; Section 3.6 shows the results or robustness checks; and Section 3.7 concludes the chapter.

3.2 Literature Review

Roy (1951) proposes a model which suggests that migration decisions are not random but imply self-selection. He emphasizes the selectivity of migration, suggesting that people with higher-than-average earnings in both their home country and in another country are more likely to immigrate when the earnings distribution in the home country is more dispersed. Sjaastad (1962) proposes a decision-making model based on cost-benefit analysis, under the assumption that migration is a kind of human capital investment. His paper suggests that migration can be viewed as an equilibrating mechanism in an economy, which promotes efficient resource allocation and facilitates individual human capital production. Wolpert (1965) posits that people make migration decision by comparing

utility at different locations, subject to imperfect information. Therefore, people's migration decisions may not be optimal. Harris and Todaro (1970) relaxes the human capital assumption by introducing imperfect labor markets in studying the migration from rural to urban areas. In their model, differences in the expected unemployment rate and income between rural and urban areas drive people to migrate.

Mincer (1978), Stark and Bloom (1985), Cobb-Clark (1990), and Borjas and Bronars (1991) argue that migration decisions are made by families or among groups, rather than by individuals. Instead of maximizing personal utility, people make migration decisions to maximize their family or group income. Carrington et al. (1996), and Bauer and Zimmermann (1995, 1997) introduce dynamic network models, which stressed the importance of networks within certain groups that help reduce migration costs.

Gravity variables such as distance between host and destination locations, and the population of host and destination locations are important factors in the migration decision (Greenwood, 1985, 1997; Van der Gaag and Wissen, 2003; Adrienko and Guriev, 2004; Cushing and Poot, 2004). Longer distance deters migration, but how population affects migration is empirically not clear (Hanson and Spilimbergo, 1996; Anjomani, 2002). Besides unemployment rate and average income (Daveri and Faini, 1999; Young, 2013; Ederveen and Bardsley, 2003), other aggregate factors, including economic variables, such as GDP and living cost (Abbott & Schmid, 1975; Kyung, 1992; Greenwood, 1997; Parikh and Van Leuvensteijn, 2003) and environmental variables, such as social amenities (Black et al., 2000; Chen and Rosenthal, 2008) and natural environment (Graves, 1979; Mueser and Graves, 1995; Adrienko and Guriev, 2004) also affect the migration decision.

Schwartz (1976) examines differences in migration behaviors for various age and education groups in terms of their responsiveness to available individual opportunities. As educational level increases, people are more likely to migrate; but the probability of migration decreases more quickly with age for more highly educated people. Other studies also stress the importance of individual characteristics in the migration decision (Greenwood, 1975, 1985, 1997; Cadwallader, 1992; Plane and Bitter, 1997; Cushing and Poot, 2004). People are less likely to migrate when they are older (Sjaastad, 1962; Schultz, 1961; Champion and Fotheringham, 1998). Some studies suggest that men are more likely to migrate than women, especially when migration is for employment (Faggian, McCann and Sheppard, 2007), whereas women are more likely to migrate than men for family-related reasons (Boyle and Halfacree, 1995). Important life events, such as employment, unemployment, marriage, retirement also play important roles in the migration decision (Rogers et al., 1978; Da Vanzo, 1978; Graves and Linneman, 1979; Plane and Heins, 2003).

A frequently mentioned personal characteristic in the migration literature is education. Empirical studies suggest that level of education is positively associated with probability of migration (Schlottmann and Herzog, 1981; Nakosteen and Zimmer, 1980, 1982; Da Vanzo, 1983). According to the summary in Haapanen and Böckerman (2013), there are several reasons for the association between education and migration. The first one, echoing Schwartz (1976), is related to earnings differentials across areas. Skilled workers have higher expected return from migration and more transferable human capital, and thus are more likely to migrate to areas providing greater average earnings (Levy and Wadycki, 1974; Wozniak, 2010). In addition, job opportunities for skilled workers are usually more

dispersed across areas (Schwartz, 1973), which drives skilled workers to migrate. Also, people with more education may be better in obtaining information on the origin and destination (Greenwood, 1997). Finally, the psychic costs for skilled workers to leave a hometown may be smaller than those for unskilled workers (Schwartz, 1973; Newbold, 1998).

However, a problem associated with identifying the causal relationship between an individual's education and the decision to migrate is that they may both be affected by unobserved individual characteristics, which can result in biased estimates of the causal impact. One approach to solve this problem is using experimental or quasi-experimental methods. Malamud and Wozniak (2012) use the draft-avoidance behavior in the Vietnam War period as an instrument for education and find a positive causal impact of education on the probability that people live outside of their birth states. Machin et al. (2012) use the reform of school instituted in 1959 in Norway as an instrument and find that education increased the probability of migration. In contrast, McHenry (2013) employs changes in the age of compulsory schooling in the U.S as an instrument for those with relatively little education and reports a negative causal impact of education on internal migration.

Some research focuses on the migration decision and the migration patterns of college-bound students and college graduates. For the study of migration of college-bound student, Tuckman (1970) finds that the average price charged by colleges, the number of public colleges, the amount of college financial aid, as well as the average income within the home state jointly explain about 64% of the variation in the proportion of students leaving their home state to attend college. Fenske (1972) finds that interstate migration declined but intrastate migration increased for college freshmen in the U.S in 1966-1969, and he

attributes the observed patterns to two national developments in higher education: erection of barriers by many states to discourage out-of-state students at state institutions, and the rapid increase in the number of public junior or community colleges, with increasing enrollments. Other empirical studies suggest that states with more educational choices and higher quality educational services, lower tuition, merit scholarship programs, higher student grants and more generous fellowships, and lower average income are more likely to retain college-bound freshmen (Steahr and Lowe, 1975; Mak and Moncur, 2003; Crook and Boyle, 2011). College-bound students consider the size of an institution, class size, difficulty of college entrance, quality of college research, and athletic programs when they make their decisions of whether to migrate outside the state to attend college (Mixon and Hsing, 1994a). Also, students with better academic performance are more concerned with the characteristics of the institution and less with the location (Faggian and Franklin, 2014).

Looking at migration decisions for college graduates, Yousefi and Rives (1987) suggest that those who are younger, male, who graduated from a high school outside of the state, or who majored in engineering are more likely to leave the state. Those who majored in agriculture, veterinary medicine, or education, or those who are looking for a job at graduation are less likely to leave the state. Kodrzycki (2001) and Tornatzky et al. (2001) both find that if a person graduate from a high school located in the same state of his/her college, the person is less likely to migrate to another state after graduating from college. Economic conditions and amenities in the state of graduation and outside states are also important in determining the migration behaviors for college graduates (Kodrzycki, 2001; Ishitani, 2010, Gottlieb & Joseph, 2006; Venhorst et al., 2011; Gottlieb & Joseph, 2006).

Although there are many studies on internal migration in China, there is not much research on the migration of college-bound students and college graduates in the Chinese context. Liu et al. (2017) find that whether university entrants migrated to attend college is mainly determined by the size of the university, while labor market differences, the distribution of universities, and cost of living are less important. For the migration of college graduates, Yue (2005) find that a critical determinant is whether the initial earnings for college graduates employed in an outside province are greater than the initial earnings they would receive if they stayed in the same province. Also, having a rural or a non-local Hukou³⁵ seems to deter high quality graduates from taking high-earning occupations. Li, Liu & Guo (2009) suggest that college graduates who migrated for employment are more likely go to big metropolitan cities in the eastern provinces than to central or western provinces. Yue (2010) find that most college graduates stay in the provinces where they graduated from. College graduates who are men, or those who graduated from better quality high schools, or those who grew up in rural areas are more likely to migrate. Also, GDP per capita in the destination has a positive effect on migration. Liu and Shen (2014) show that both skilled and less skilled workers tend to migrate from economically less developed inland areas to more developed coastal areas. They attribute the similarity in migration patterns between the skilled and unskilled to both well-established labor-intensive industries and fast-growing knowledge-based industries in the coastal provinces. Employment opportunities and wage differentials play a dominant role in attracting skilled labor, whereas the impact of amenities is less clear. Liu et al. (2017) confirm that skilled

³⁵ "Hukou" is a system of household registration in mainland China. It records an individual's registered address and the types of household (agricultural or non-agricultural). It also records basic demographic information of household members.

workers have a strong tendency to stay in the same province where they obtained their higher educational levels. The migration of college graduates is mainly driven by regional differences in wages but not the differences in unemployment rates. Fu and Gabriel (2012) suggest that concentration of educated populations in a region has a significantly positive impact on the migration odds for people at the highest educational levels. Zheng et al. (2020) study the migration pattern for skilled floating population³⁶ in China using data from a job-recruiting platform. They find that the provincial capitals attract skilled floating population within the same province or from surrounding provinces. The first-tier cities attract skilled migrants nationwide. Intra-provincial skilled migrants are more responsive to employment opportunities, and inter-provincial skilled migrants are more responsive to amenities. Skilled migrants from more prestigious universities are more flexible in choosing migration destinations. Shen and Liu (2016) use a random sample from a census data in China and study the migration patterns for skilled labor and less-skilled labor. They find that skilled people are more likely to leave less populous provinces and are more likely to leave origin provinces when the total number of university graduates increases. Also, skilled labor is more sensitive to urban wage level and less sensitive to unemployment in origin province when making migration decision.

There is a gap in the previous literature on the migration in China as very few studies have discussed the direct impact of educational policies on migration decisions. In this chapter, we analyze the impact of the college expansion on migration behaviors for college-bound students, college graduates, and non-college graduates. This study contributes to the

³⁶ The floating population includes those whose registered addresses (Hukous) are in a different county than from residential address (different in county level) for at least 6 months.

literature by studying the direct effects of the college expansion on the migration decision of those during the ages of college attendance, and in the period after graduation for both college graduates and others. We also analyze the economic channels for the observed impacts.

This chapter also extends the second chapter in this dissertation, “Impact of the College Expansion on Skilled Labor Market in China” in studying the effect of the college expansion on the labor market outcomes (income, unemployment, usage of different skills) in the short run. In this chapter, we study the impact of college expansion on migration decision after graduation and confirm that the college expansion generates competition effect on migration for college graduates after graduation. The results in this chapter partly explain the downward bias in the estimated effect of college expansion on labor market outcomes in the second chapter.

3.3 Data and Background

We use the 2012 Chinese Labor Dynamics Survey (CLDS) survey to conduct our study. The CLDS is the first national longitudinal study of the Chinese labor force, launched by the Sun Yat-Sen University. It is a nationally representative biannual rotation panel survey, which was initiated in 2012 and covers 29 provinces in mainland China (Tibet and Hainan are excluded). The sampling method of this survey uses probability-proportional-to-size sampling (PPS) of households, and the targeted respondents of the survey are individuals aged 15-65 at the time of the survey. The survey is suitable for studying migration behavior in different ages since it asks people details of their first five migratory moves, including when they migrated, where they migrated from and migrated to, and why they migrated. It also obtains basic demographic information, including

education, Hukou status, Hukou change history, work history, and parents' information. By summarizing the micro-level statistics from the survey data and comparing them with the macro-level statistics of migration, we can obtain an overall understanding of migration patterns in China.

3.3.1 Migration Patterns in China in Recent Decades

The volume of migration has surged over the period of Reform and Opening beginning in the 1980s due to the development of the economy and relaxation of migration restrictions. The Chinese government defines migration in a different way from that of most migration studies. The official measure of migration is related to people's registered address (Hukou). One measure of migration counts a person as a migrant if the person lives in a different address than his/her Hukou address for at least 6 months. A more frequently measure of migration is based on what is called the "floating population", which includes people who live in different addresses than their Hukou addresses for at least 6 months, but it does not count any within-county address discrepancy. Figure 3.1 shows the volume of the floating population in China from 1982 to 2015, as well as the percentage of this group in the whole population, and the regional distribution among the migration destinations. We can see that the floating population increases nearly 40 times from 1982 to 2015, and the floating population fraction increases from 0.7% to 18.2%. As for the destination of migration, eastern provinces always attract the most migrants, accounting for 40%-60% of all migration destinations. Western provinces attract about 20%-30% migration destination, and central provinces attract about 15%-25% migration destinations.

In Figure 3.2, we show the educational distribution of the floating population in China using the results from two recent Chinese censuses. As can be seen from the figure, the

largest educational categories of migrants in China contain those who have middle high school diplomas (9 years of education) or senior high school diplomas (12 years of education). Migrants in 2010 are more educated than migrants in 2000, since the proportion of college graduates and above increased by 4.8 percentage points. This is due to the college expansion in China since 1999, which greatly increased the number of college graduates.

In Figure 3.3, we plot the inter-province and intra-province migration rates and their 95% confidence intervals for ages 17-20, 21-25, and 26-34 for non-college graduates and college graduates in 1984-1991, 1992-1998, and 1999-2012. The migration rate is calculated using the number of observations where the respondent stated they had changed their place of residence either to a different province (inter-provincial), or within a province (intra-provincial) in an age range (e.g., migrated in age 17-20) during a time period (e.g., migrated in 1999-2012), divided by the number of observations for which the age range occurs within a specified time period (e.g., observations who were at least 17 in 1999 and who were at most 20 in 2012).

As can be seen from the graph, the inter-provincial migration rates for both college graduates and non-college graduates increase over time. For non-college graduates, the rate in every age group more than doubles from 1984 to 2012. The inter-provincial migration rate for non-college graduates increases relative to that for college graduates for ages 17-20 across time, and it is 4 percentage points greater than the migration rate for college graduates in 1999-2012. The increasing gap in inter-province migration between college graduates and non-college graduates in younger ages may be explained by the relaxing of Hukou restrictions in the late 1980s and China's entry into the WTO, which provided

opportunities for non-college graduates to work in coastal provinces. The inter-provincial migration rate increases for college graduates in each age group, although the growth rate is smaller for them. The increase in the migration rate for college graduates for ages 21-25 triples in 1999-2012 compared to 1992-1998. In addition to the general changes in the economy noted above, another possible reason may be the college expansion starting in 1999, which greatly increases the skilled labor supply and thus changed their migration behavior.

The intra-provincial migration rates also increase greatly from 1984 to 2012 for both college graduates and non-college graduates. The intra-provincial migration rates are higher for college graduates than non-college graduates for ages 17-20 in all time periods. But due to relatively small sample sizes of college graduates in early periods, the migration rates are estimated with great uncertainty for ages 21-25 and ages 26-34 in 1984-1991, and for ages 26-34 in 1992-1998.

Figure 3.4 shows the distribution of migration reasons provided by the CLDS respondents. We can see that, in the age range 17-20, in all periods more than 60% of the college-graduate migrants migrated to obtain education, but migration for work accounted for at least 75% of migrants who didn't have college degrees. Considering the migration of college graduates when they were 17-20 years old, in 1992-1998 they were less likely to migrate for education than in 1999-2012; and compared to those in 1984-1991, they are less likely to migrate to join the army. The proportions of migrants who state work as the reason for migration in ages 21-25 is similar for non-college graduates and college graduates in 1984-1991 and 1999-2012, but higher for non-college graduates in 1992-1998. That some people choose to take part-time college courses after they start working is a

possible explanation for the increase of “education” portion in 1992-1998. After 1999, since college admissions expanded greatly, it is much more likely for a person to attend college in the age range 17-20, rather than later, when they already have a job. Due to small sample sizes for college graduates in early periods, we are not able to plot migration rates for them for ages 26-34 in 1984-1998. But based on what we observe in 1999-2012, although non-college graduates in the age range 26-34 are likely to migrate for work, this reason accounts for a very small portion of college-graduate-migrants for ages 26-34. They are more likely to migrate for family or other reasons in this age range.

As noted in the previous paragraphs, the household registration system, Hukou, plays a very important part in people’s lives, especially in migration decision making. Local governments determine the public services a person can receive based on whether the person has a local Hukou, and whether the Hukou type is agricultural or non-agricultural. For those born elsewhere who have only a high school diploma, it is not easy to obtain Hukou registration in a large city, like Beijing or Shanghai. Also, there are some restrictions if a person wants to change his/her Hukou type. Figure 3.5 plots the migration rates and the Hukou change rates for ages 17-20, 21-25, and 26-34 for non-college graduates and college graduates in 1984-1991, 1992-1998, and 1999-2012. The patterns for migration rates are similar to what we observe in Table 3 in that they all increase dramatically for both college graduates and non-college graduates from 1984 to 2012, and the gaps of migration rates between non-college graduates and college graduates become greater over time. Hukou change rates are higher for college graduates than for non-college graduates in all age groups and in all periods since the requirements of changing Hukou to a more economically developed region are highly related to education. The traditional ways

to change Hukou status from rural to urban or to a large city is through enrolling in a college, joining the army, and marriage. For example, students enrolled in a college located in a large city can change Hukou to that city and receive public services there. Upon graduation, if they obtain employment in the city, they can retain that Hukou designation.

However, Hukou change rates do not increase much from 1984 to 2012 for young people. For older people (ages 26-34), the Hukou change rate increases greatly across time, suggesting that the Hukou relaxation likely benefited older people most. Although non-college graduates in the age range 17-20 are more likely to migrate, they are much less likely to change their Hukou than college graduates. This suggests that non-college graduates are more likely to migrate without changing their Hukou, which is because the requirements for changing a Hukou are more challenging for them. In fact, those admitted to college often are permitted to obtain a new Hukou designation, and those who obtain government or higher-level positions are generally able to obtain such a change (The Regulations on Household Registration of the People's Republic of China³⁷). Those who are not able to modify their Hukou are less likely to migrate permanently since Hukou is highly associated with the public services a person can access in the registered location.

More evidence on the relationship between migration and Hukou change can be seen in Figure 3.6, which plots the tabulation of migration and Hukou change for different age groups in 1984-1991, 1992-1998, and 1999-2012. This figure suggests that college graduates are more likely to change Hukou without migration than non-college graduates

³⁷ URL: http://www.npc.gov.cn/wxzl/gongbao/2000-12/10/content_5004332.htm (in Chinese); See <https://baike.baidu.com/item/%E6%88%B7%E5%8F%A3%E8%BF%81%E7%A7%BB> (in Chinese) for more detailed interpretations about the regulations.

for all age ranges in all periods. Changing Hukou but without migration is possible since a person may migrate at an earlier age without changing Hukou and satisfy the requirements to change Hukou later. For example, a person may migrate before 17 with his/her parents and enroll in a local college after age 17. Our data supports this explanation. We also observe that non-college graduates are more likely to migrate without changing Hukou than college graduates, especially for those who were in the age range 17-20 in 1999-2012. This pattern also supports the view that non-college migrants are more likely to be temporary migrants rather than permanent migrants.

Figure 3.7 shows the distribution of destination region type for people who change their Hukou by age range and by education for 1984-1991, 1992-1998, and 1999-2012. Due to small sample sizes in early periods, we only plot Hukou change reasons for ages 17-20 and ages 21-25. We can see from this graph that there is a large difference between the college-graduate and non-college-graduate migrants in terms of type of destination region for Hukou change.³⁸ In all periods and both age ranges, non-college migrants are much more likely to change their Hukou addresses to villages and towns, which are the least urbanized areas in China. College migrants are more likely to change their Hukou addresses to cities and municipalities, which are the more urbanized areas. We also see that both non-college graduates and college graduates in the age range 17-20 are more likely to change their Hukou to provincial capitals in 1999-2012 than before. The reason for this increase may differ for college graduates and non-college graduates. For college graduates, such an increase may be due to the nationwide college expansion that increases the college

³⁸ The regions are ordered by their degree of urbanization (from least to the greatest): village and town, suburb, city, provincial capital, and municipality.

admissions in each province; since the higher education resources are concentrated in provincial capitals, they will attract college attendees (most colleges allow students to change Hukou). For non-college graduates, the increase may be due to the relaxation of Hukou restrictions in cities and more rapid economic growth in provincial capitals and in municipalities after 1999.

Figure 3.8 plots the reasons for changing Hukou by age group and by education in 1984-1991, 1992-1998, and 1999-2012. From the figure we can conclude that the top reason for non-college graduates to change their Hukou is family-related reasons in all age groups and in all time periods. College graduates who changed their Hukou in 17-20 are most likely to state education as the reason, and the proportion is even greater in more recent periods, reaching 95% of the Hukou change reasons in 1999-2012. This is most likely due to the college expansion beginning in 1999. Also, college graduates are more likely to change Hukou for work than non-college graduates for ages 21-25, although they are the same or less likely to migrate for work-related reasons (see Figure 3.4) than non-college graduates for ages 21-25. These patterns also suggest that, in China, less-educated migrants are more likely to be temporary migrants, who migrate for work but do not ask for access to basic local welfare services offered to Hukou members. In contrast, more-educated migrants are more likely to be permanent ones.

3.3.2 College Admission Planning and the College Expansion beginning in 1999

In 1999, the Chinese government announced a decision to accelerate the expansion of higher education in order to increase human capital accumulation, stimulate domestic demand, and improve employment options for young people and workers laid off from state owned enterprises. In the Fifteenth National People's Congress, the Ministry of

Education of the PRC published guidance for higher education development in China for the 21st century. It set a goal for the higher education enrollment rate to reach 11% by 2000 and more broadly to dramatically increase the number of college-educated people in the 21st century. The main policy associated with this guidance is increasing the college admissions in all provinces of mainland China.

Figure 3.9 shows the number of college admissions and their regional distribution in China from 1992 to 2012. From 1993 to 1998, the number of admissions stayed around 2 million. In 1999, the admissions increased by 48%, and it kept increasing more than 10% every year to 2012. The number of college admissions increased by nearly 4 times from 1998 to 2012. Also, the proportion of college admissions in the eastern provinces account 55% before 1999, but it gradually decreases to 48% by 2012. In contrast, the proportion of college admissions in central provinces increase by 3 percentage points and the admissions in western provinces increase by 4 percentage points. The change in the regional distribution of college admissions suggests that the college expansion reform attempted to balance higher education resources across regions.

Figure 3.10 plots the admissions by region (eastern, central, western) for the nine types of universities, which are listed by their relative admission sizes: STEM and trade-related colleges (finance and economics, language, law and politics), comprehensive colleges, 2- or 3-year vocational colleges, and other colleges (medical, teacher, agricultural and forestry). We can see from these graphs that the admissions increase in eastern provinces are less than the admissions increase in the central and western provinces for all types of universities. These graphs suggest that the college expansion allocated more college admissions to the regions that lacked higher educational resources before the expansion

than those with more educational resources. Also, since the percentage of population living in eastern provinces increased from 40% to 44% from 1998 to 2012, it confirms that the greater increase of college admissions in central and western provinces is not corresponding to the greater increase in their overall population. This pattern may decrease the inter-provincial migration for people who attend a college, since it provides college opportunities across provinces more in line with their relative populations. It may also attract people from eastern provinces to attend a college in central or western provinces and therefore increase the inter-provincial migration.

3.3.3 Descriptive Statistics from Chinese Labor Dynamics Survey

Table 3.1 lists summary statistics for our sample from the Chinese Labor Dynamics Survey (2012) by birth cohort. The table summarizes the mean age, schooling, proportion male in all cohorts, and birth cohort (1952-1962, 1963-1972, and 1973-1982). It also summarizes the migration rate (both inter- and intra-province), reason of first migration, and first migration destination region under age 30. “Migration” is defined as the movement from a township to areas outside of the township (for rural residents), or from one sub-district to areas outside of the sub-district³⁹ (for urban residents), and a stay in the destination for at least 6 months. As we expect, the average years of schooling and the migration rate are greater for younger migrants. People born in 1973-1982 are about 4 times more likely to migrate than people born in 1952-1962. Looking at the reasons for migration, the proportion listing political (the Down to the Countryside Movement,

³⁹ Sub-district is a division under the district level, which is the division under the city level. According to the Constitution of the PRC, a sub-district must have at least 50,000 residents.

assignment to support construction in the border areas⁴⁰) or family related reasons decreases across birth cohort, while people born in later cohorts are more likely to migrate for work or to pursue education. The average ages at first migration are about the same across birth cohorts, and migrants in later generations are more likely to migrate to eastern or western provinces but less likely to migrate to central provinces.

To better understand the pattern of different types of migration, in Table 3.2 we present the mean and standard deviation of age and education at the time of migration, gender, and Hukou status at birth by reason for migration (family, work, and education) under age 30, as well as the destination region of migration under age 30. The table is also divided by migration cohort into the pre-college expansion period (1978-1998) and the post-college expansion period (1999-2012).

Table 3.2 shows that those who migrated in more recent cohorts were older and more educated at the time migration, more likely to be female, and more likely to have grown up in rural areas than the migrants who migrated before 1999. Those who indicate they migrated for education are in general younger, more educated, and more likely to have grown up in urban areas than those who migrated for other reasons. People who migrated for work are more likely have grown up in rural areas. For the regional distribution, we see that people who migrated for work are more likely to choose eastern provinces than those who migrated for other reasons, and the proportion of migration to the eastern provinces increase after 1999 for this group of migrants while it decreases for the other two groups.

⁴⁰ The Down to the Countryside Movement during the Cultural Revolution in China sent 17 million urban young people to rural areas. Assignment to support construction of the border areas occurred under a series of policies which aimed to bring people from developed regions to less developed regions on the border.

This may be due to the huge economic growth and therefore increasing opportunity of employment in the eastern provinces since China's entry into the WTO. The increasing proportion of migration for education to central provinces may be related to the college expansion policy beginning in 1999, since college growth was greatest in less developed regions.

3.4 Models

3.4.1 Linear Probability Model

We first start our analyses by applying a simple choice model. Assume that a person k 's utility for living in a location i is U_i , which follows the form:

$$U_i^k = u_i + \varepsilon_i^k \quad (3.1)$$

where u_i is the utility of living in i , which is assumed to be the same for all individuals. ε_i^k is an error term that varies at both the individual and location level. Let M_{ij}^k denote the decision for a person k to migrate from i to j , and c_{ij} the cost of migration from i to j , which is assumed to be the same for each person. Then, the person will choose to migrate only when: $U_j^k - U_i^k - c_{ij} \geq U_l^k - U_i^k - c_{il}$ for all l . Therefore, the probability for person k to migrate can be written as:

$$Pr(M_{ij}^k = 1) = Pr((u_j - u_i - c_{ij}) - (u_l - u_i - c_{il}) + \varepsilon_j^k \geq \varepsilon_l^k, \text{ for all } l) \quad (3.2)$$

Following McFadden (1974), we assume that ε_i^k follows an extreme value type I distribution for all i and k . Therefore, the probability of migration from i to j for person k is:

$$Pr(M_{ij}^k = 1) = \frac{e^{(u_j - u_i - c_{ij})}}{\sum_{all\ l} e^{(u_l - u_i - c_{il})}} \quad (3.3)$$

Also based on equation (3.2), the aggregate migration flow between i and j is:

$$M_{ij} = \sum_k Pr(M_{ij}^k = 1) = POP_i \frac{e^{(u_j - u_i - c_{ij})}}{\sum_{all\ l} e^{(u_l - u_i - c_{il})}} = POP_i D_i' D_j d_{ij}^\gamma \quad (3.4)$$

where D_i is an origin-specific vector, D_j is a destination-specific vector, and POP_i is the population of location i . We assume that the migration cost from i to j is mainly the transportation cost which depends on the distance between province i and province j , d_{ij} , so, $c_{ij} = \alpha + \gamma \ln d_{ij}$. Thus, the probability that person k migrates from i to any province outside of i (interprovincial migration) is:

$$Pr(M_i^k = 1) = \sum_{j \neq i} \frac{M_{ij}}{POP_i} = D_i' \sum_{j \neq i} D_j d_{ij}^\gamma \quad (3.5)$$

Linearizing this expression and introducing time variation and individual differences, we can write the migration choice as a linear probability model of the form:

$$Pr(M_{it}^k = 1) = \beta_1 D_{it} + \beta_2 D_{-it} + \beta_3 X^k + \delta_i + \delta_t + \varepsilon_{it}^k \quad (3.6)$$

where M_{it}^k is a dummy for whether the person k migrated outside of province i in year t . D_{it} is a vector of characteristics of origin province i (which is the province the person k came from) in year t . We include our variable of interest, college expansion (measured as the number of provincial college admissions), and gravity (population) and labor market variables (average wages for skilled-labor intensive industries, average wages for unskilled-labor intensive industries, unemployment rate), and personal characteristics (gender, Hukou type (agricultural or non-agricultural) at birth, years of father's schooling).

The descriptive statistics of these variables are listed in APTable 3.1.

Each of the variables in D_{it} is normalized by subtracting its mean value. D_{-it} is an “other provinces” vector, which is a weighted vector based on all provinces other than i . It contains the same variables as those in D_{it} , measured at each of provinces other than i , adjusting by the distance between the central point of the origin province i and the central point of the province j , i.e.

$$D_{-it} = \sum_{j \neq i} D_{jt} w_{ij} \quad (3.7)$$

The weight between i and j , w_{ij} it was obtained from a regression of migration flow from i to j on the distance between i and j , fixed effects for origin province i , and fixed effects for destination province j :

$$\ln(\sum_k M_{ij}^k) = \beta_1 D_i + \beta_2 D_j + \gamma \ln d_{ij} + \varepsilon_{ij} \quad (3.8)$$

$$w_{ij} = f(x) = \begin{cases} \ln \widehat{\sum_k M_{ij}^k} / \sum_j (\ln \widehat{\sum_k M_{ij}^k}), & j \neq i \\ 0, & j = i \end{cases} \quad (3.9)$$

where $\sum_k M_{ij}^k$ is the total migration flow from i to j , and $\ln \widehat{\sum_k M_{ij}^k}$ is the predicted value based on equation (3.8). We obtained the migration flow data from the 2012 China Migrant Dynamic Survey, which is a nationally representative survey of migrants in 31 provinces of mainland China with sample size of 198,795. D_i and D_j are fixed effects of migration origin province i and migration destination province j , respectively. d_{ij} is the distance between the center of province i to the center of province j (km), and we assume $d_{ij} = 1$ for $j = i$. The weight between province i and j is therefore w_{ij} , which is the ratio of the log of predicted migration flow between i and j to the sum of the log of predicted migration flow from i to each of 31 provinces in China for $j \neq i$. We assign $w_{ij} = 0$ if $j = i$, and, by

construction, $\sum_j w_{ij} = 1$. X_i in equation (3.6) is a vector of personal characteristics, including gender, age, age squared, father's schooling, and Hukou status when the person was born. δ_p is the province fixed effects, δ_t is the year fixed effects, and ε_{it}^k is an error term.

3.4.2 Conditional Logit Model

We estimate the conditional Logit model of migration between region i (origin region) and region j (destination region) for person k . We categorize China's provinces into three regions: eastern, central, and western, and each individual chooses to move to (or stay in) one of these regions. Here, M_{ij}^k in equation (3.3) is a dummy that denotes a change of residential region from region i (origin region) to region j (if the individual does not move or moves within a region, then destination region i is equal to j) for person k in year t . Incorporating the individual characteristics into the model, we update the equation to:

$$Pr(M_{ij,t}^k = 1) = \begin{cases} \frac{1}{1 + \sum_{l=2}^3 e^{(\beta_{1l}X^k + \beta_{2l}A_{il} + \beta_{3l}D_{lt} + \delta_l + \varepsilon_{lt}^k)}}, & \text{if } j = 1 \\ \frac{e^{(\beta_{1j}X^k + \beta_{2j}A_{ij} + \beta_{3j}D_{jt} + \delta_j + \varepsilon_{jt}^k)}}{1 + \sum_{l=2}^3 e^{(\beta_{1l}X^k + \beta_{2l}A_{il} + \beta_{3l}D_{lt} + \delta_l + \varepsilon_{lt}^k)}}, & \text{if } j > 1 \end{cases} \quad (3.10)$$

M_{ijt}^k is a dummy for whether the person k from region i migrated to region j in year t , where $M_{iit}^k = 1$ implies that the individual did not move from region i . X^k is the vector of personal characteristics, including gender, age, age squared, father's schooling, Hukou type when the person was born, and a dummy for whether the person migrated before. Personal characteristics are region-specific, so the coefficients differ by region. A_{ij} is a vector including variables that measure the relationship between origin region and destination region, which includes the average distance (see below) from the region i to region j ,

$Dist_{i,j}$, and a “Home” dummy for whether the destination region is equal to the origin region, $Home_{i,j}$. $Dist_{i,j} = \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} d_{k,l} / (N_i * N_j)$, where k is the origin province within region i , and l is a destination province within region j . $d_{k,l}$ is the distance between the centers of provinces k and l , N_i is the number of provinces within a region i , and N_j is the number of provinces within a region j . D_{jt} is a time-variant vector for destination j which includes average wage for skilled labor-intensive industries ($SWAGE_{jt}$), average wage for unskilled labor-intensive industries ($UWAGE_{jt}$), average unemployment rate (UR_{jt}), and population (POP_{jt}), in the region j in year t . δ_j is the destination region fixed effects, and ε_{jt}^k is the error term.

3.5 Results

3.5.1 Results Using Linear Probability Model

3.5.1.1 Enrollment Effect

College expansion may affect migration behavior when a person makes his/her decision of whether to attend college. Increasing college enrollment opportunities in the local province may cause people who plan to attend college to stay, but greater enrollment opportunities in outside provinces may cause increases in movement to other provinces.

We analyze the migration decision at age 17-20, which is the typical age range when a person chooses whether to attend college or start working. In equation (3.6), t is thus the calendar year at age 16 for each individual k (i.e., all of the aggregate variables are measured for the calendar year when person k is age 16). The estimation results for the migration decision at age 17-20 are shown in Table 3.3. In column (1), we only include our variable of interest, local college admission (AD) and the average college admissions in

other provinces (otherAD), as well as gravity variables (POP, otherPOP) and personal characteristics (gender, father's schooling, and Hukou type at birth), but without including individual college enrollment or labor market variables. In column (2), we include labor market variables, average unskilled labor wage and unemployment rate). In column (3), we add the college degree attainment dummy, which measures whether the person has a college degree at the time of survey. We use the Col as a measure of individual college enrollment to see if the impact of aggregate college admissions on individual migration is mediated through individual college enrollment. Finally, in column (4), we add the interaction terms between aggregate college admissions and individual college degree attainment (as a proxy for individual college enrollment), as well as the interaction terms between labor market variables and college degree attainment to see whether the effects of these measures on migration differ for those who attended college as compared to others. The standard errors are clustered in province level. However, since the data only cover 29 provinces in mainland China, which is smaller than the standard requirement for the number of clusters in a regression (rule of 42). We applied the wild bootstrap test proposed in Cameron et al. (2008, 2010) to correct the bias of standard errors due to small number of clusters.

We use college admissions in each province as a measure of the degree of college expansion. As we can observe in Table 3.3, the college expansion in the local province decreases the probability of migration by about 2 percentage points per 100,000 increase in college admissions. Although the estimate of the effect for the college admissions in outside provinces is positive, the coefficient is statistically insignificant. Adding labor market variables in column (2) and adding a college degree attainment dummy in column

(3) do not change the estimated results very much. Interestingly, we also find that having a college degree has a direct effect, decreasing the probability of inter-provincial migration by 4 percentage points. This result is consistent with the migration pattern in China (see Figure 3.3). An explanation is that China joined the WTO in 2001, which created many job opportunities in coastal provinces for less-skilled workers and causing inter-provincial migration probability to increase greatly for less-skilled workers.

In column (4), we study the differential impacts of the college expansion on migration for college attendees and others by adding interactions between college degree attainment, which is a proxy for individual college enrollment, and college admissions and labor market factors. We find that as the distance-weighted average college admissions (see the weighting formula in equation (3.7)) in non-local provinces increase by 100,000, the probability of inter-provincial migration for people who attend a college increase by 11 percentage points, relative to non-college graduates. This suggests that the college expansion in other provinces attracts people who plan to attend a college to migrate inter-provincially. However, although the negative impact of local admissions does not change when adding these interactions, we do not find any statistically significant evidence to suggest that the local admissions cause college graduates to be more likely to stay than non-college graduates. In other words, an increase in local admissions makes both those who plan to attend college and those who do not plan to attend college more likely to stay. After controlling for the interactions, the estimated coefficient for college attainment (at the mean values of other macro-level variables) become positive but is not statistically significant, suggesting that the higher probability of migration for non-college graduates can be mostly explained by better economic and labor market conditions.

We conclude that there is both an individual college enrollment effect, and an aggregate college enrollment effect on inter-provincial migration due to the college expansion. Specifically, as the average number of college admissions in outside provinces increase, those who later obtained a college degree are more likely to migrate to another province during ages 17-20. As local aggregate college admissions increase, both non-college graduates and college graduates are less likely to migrate to another province. College expansion itself provides work opportunities, such as campus construction and maintenance, and boosts business related to college education, so even a person who does not attend college may be induced to stay in an area with college expansion—it is hard to distinguish the enrollment effect from this general economic effect using limited information. Also, people who later obtained a college degree are more likely to migrate as the average wage of skilled-labor-intensive industries in outside provinces increases.

3.5.1.2 Competition Effect

The second channel through which college expansion may affect s migration behavior is increasing the competition of young skilled labor when they enter the job market. An increasing number of new college graduates in a given cohort year within a province may cause new college graduates to migrate to other provinces to avoid competition in the local job market.

Table 3.4 reports the estimated results for people’s migration decisions in the 3 years after graduating from their highest level of education. In this analysis, we use college admissions when the person was age 16 in the province where a person graduated from his/her highest level of education. We apply a similar model as shown in Table 3.3.

We observe that when not controlling for the interactions between college degree with admissions and labor market variables (columns (1)-(3)), the results suggest that an increase in the college graduates of other provinces induces migration after graduation. For every 100,000 increase in the average admissions of other provinces when a person was 16, it increases the probability of his/her migration after graduation by about 6 percentage points. However, we do not find a statistically significant effect of the increase in the number of local college graduates on the migration after graduation.

After adding interactions in column (4) in Table 3.4, we see that college graduates are more likely to migrate inter-provincially after graduation in response to increases of college admissions in their local province at age 16, relative to non-college graduates. For every 100,000 increase in the average number of college graduates in the local province, the inter-provincial migration probability for college graduates increases by about 1 percentage point. In contrast, as average number of college graduates in the local province increases by 100,000, the probability of migration decreases for non-college graduates by 3 percentage points. An explanation is related to the competition effect that if a province admitted many college students, it will be hard for college graduates to find a job after graduation due to greater supply of skilled workers within a province. Therefore, college graduates would migrate to a province with less competition after graduation. Non-college graduates may respond similarly if they are complementary to skilled workers, so more college graduates in the local province induce them to stay. We do not find any significant impact of the college admissions in other provinces on either college graduates or non-college graduates.

3.5.2 Results Using Conditional Logit Model

We use the conditional Logit model to study if college admissions affect migration to specific regions. Tables 3.5 and 3.6 list the estimated marginal effects of admissions on migration to particular destinations for ages 17-20, and the marginal effect on migration within 3 years after graduation. As described in the Model section, macro-level variables used in the conditional Logit model are measured at the regional (eastern, central, and western) level. The base group is the eastern region. Therefore, the estimated coefficients for each macro-level variable can be interpreted as the effect of that factor on whether an individual migrates to or stays in a given region; and the coefficient for a personal characteristic can be interpreted as the effect of that characteristic on whether the person migrates to a specific region. Recall that region fixed effects are controlled in this model, so estimates of responses to regional macro-level variables are based on variation over time.

In the third column of Tables 3.5 and 3.6, we include an interaction term between number of college admissions and the home dummy, and an interaction between AD and (1-Home). We also include all of the labor market variables as mentioned in Tables 3.3 and 3.4, their interactions with the home dummy, and their interactions with the non-home dummy. Personal characteristics are controlled, too.

As can be observed from Table 3.5, although the estimated coefficients for the interaction terms between AD and college degree attainment are positive, and the estimated coefficients for college admissions are positive, they are not statistically significant. We also observe that when distance between two regions increases, which we define as the average distance from a province in one region to a province in another region (see model section), people are less likely to migrate. As distance between two regions increases, the

probability of migration decreases. Also, people are more likely to stay in their home regions. For example, the predicted probability of staying in an eastern province is 0.96, compared with the predicted probability of movement from an eastern province to a central province, 0.01, given 1,000 km distance between an eastern province and a central province; and the predicted probability of movement from an eastern province to a western province, 0.03, given 1,000 km distance between an eastern province and a western province.

We observe different effects of personal variables on migration to different regions. The results suggest that a person with a college degree or a non-agricultural Hukou is relatively more likely to migrate to the central and western provinces, relative to the migration to eastern provinces. These results may be tied to China's entry into the WTO, which increased opportunities of jobs for less-skilled labor in the eastern provinces.

Table 3.6 reports the estimated marginal effect of variables on migration to each region in the 3 years after graduation using the conditional Logit model. We observe a similar "competition effect" at the regional level. The results suggest a negative effect of the increase in new college graduates in a region on a person's migration to that region after graduation. As the number of college admission in local region increases by 100,000, the probability of staying in the origin region decreases by 7 percentage points.

Similar to Table 3.5, we observe a negative effect of distance on migration and a positive effect of home region on migration. For example, the predicted probability of staying in an eastern province is 0.89, compared with the predicted probability of movement from an eastern province to a central province, 0.03, given 1,000 km distance

between an eastern province and a central province; the predicted probability of movement from an eastern province to a western province is 0.05, given 1,000 km distance between an eastern province and a western province.

We also find different effects of personal variables on the migration to different regions after graduation from the migration decision at ages 17-20. People with college degrees are less likely to migrate to central or western provinces, relative to eastern provinces. People with an agricultural Hukou status at birth are more likely to migrate to central and western provinces.

Comparing Tables 3.5 and 3.6 with Tables 3.3 and 3.4, we find similar migration patterns at ages 17-20 and after graduation using the conditional logit model. However, the “individual enrollment effect” at ages 17-20 is not statistically significant in the regional choice model of migration. The “competition effect” after graduation we observe in inter-provincial migration also is observed for migration decisions as modeled in the regional choice model. However, the “aggregate enrollment effect” at ages 17-20 for both college attendees and non-college attendees is still observed in the regional choice migration model, suggesting that a region with more college admissions attracts all people in ages 17-20. The conditional logit results also confirm that college graduates are more responsive to labor market conditions than non-college graduates when making inter-regional migration decision after graduation.

To test the Independence of Irrelevant Alternatives (IIA) property, we drop eastern, central, and western region, respectively and test if these alternative specifications change the estimated results using the conditional Logit model. Following Hausman and

McFadden (1984), we use a constructed IIA test to see if the coefficients differ after dropping each region of out the sample:

$$\chi^2 = (\beta_s - \beta_f)'(V_s - V_f)^{-1}(\beta_s - \beta_f)$$

Where β_s is the vector of the estimated coefficients using subsample (after dropping a region), and β_f is the vector of the estimated coefficients using full sample. V_s is the covariance matrix of estimators for the subsample, and V_f is the covariance matrix of estimators for the full sample. The degrees of freedom for this test are equal to the rank of $(V_s - V_f)$. The chi square scores for the comparisons are listed below:

As can be seen from Table 3.7, the calculated chi square value of dropping each region is smaller than the 95% critical value with 17 degrees of freedom (i.e., the rank of the covariance matrix). These results suggest that the estimated coefficients when dropping either eastern, central, or western provinces change little compared with the estimated coefficients using the full sample. Therefore, the IIA property is satisfied.

3.6 Robustness Checks

3.6.1 Removing Migration to Eastern Provinces

During the period which we study (1992-2012), China experienced tremendous economic growth due the “Reform and Opening” policy begun in 1978 and following its entry into the WTO in 2001. The total value of exports and imports increased almost 22 times from 1992 to 2012.⁴¹ Accompanying increasing volume of trade, factories in the eastern provinces expanded greatly and recruited many manufacturing workers. Attracted

⁴¹ See the “Table 6-3 Total Value of Imports and Exports” in China Statistical Yearbook 2013, <http://www.stats.gov.cn/tjsj/ndsj/2013/html/Z0603e.htm>

by higher wages, people from less-developed areas migrated to the eastern provinces. This migration trend is often identified as the dominant one in internal migration after 2001 (Zheng & Yang, 2013).

Since our goal in this study is to identify the effect of college expansion in China on migration, which started at a year near China's entry into the WTO (1999), one concern is that we are not able to distinguish the impact of the increasing trade on migration from the impacts we try to identify. Therefore, we have removed all the migration to eastern provinces and redone the estimation of the previous section. By comparing these results with those reported above, we can check if our estimates are affected by not excluding this "main trend" in migration. Therefore, the first check on our estimation is to replicate analyses above removing migration to the eastern provinces. Specifically, in this robustness check, the migration dummy is re-defined as whether a person migrates to a non-eastern province where the destination province is not the original province. Equivalently, the outcome variable equals 1 when a person migrated to a non-eastern province outside of original province, and it equals to 0 otherwise. APTable 3.2 shows the estimated results of the non-eastern migration at age 17-20 and APTable 3.3 shows the results of the non-eastern migration in 3 years after graduation. In these tables, we estimate the same coefficients as shown in Table 3.3 and Table 3.4 (we do not report the regression results in the original column (3) of Tables 3.3 and 3.4).

We find that after excluding the migration to eastern province, the results still indicate similar patterns as what we summarized from Table 3.3 and Table 3.4. They confirm the pattern that local college admissions reduces the probability of migration for ages 17-20, no matter whether it is to eastern provinces or not. APTable 3.3 supports the conclusion

that the increase of local skilled worker supply (proxied by local college admissions in age 16) reduces the probability of departure for less-skilled workers. We also observe similar competition effects of local aggregate admissions on migration after graduation. However, comparing the magnitude of the estimated regression coefficients between APTable 3.2 and Table 3.3, and APTable 3.3 and Table 3.4, we find that the estimated “enrollment effect” and “competition effect” are smaller in the regression of non-eastern migration. This indicates that these effects, in part, are based on patterns of migration to eastern provinces.

3.6.2 Using Alternative Proxy for College Expansion

An additional test is to change our measure of college expansion to be the number of teachers in colleges in each province in each year. One may argue that using number of student entrants in colleges as the measure of college expansion may bias the estimation since although the level of college admissions is planned by the government, the final number of new entrants is directly affected by individual decisions. To check whether this potential problem causes bias, we use number of teachers in colleges as a measure of college expansion. The number of teachers is less likely to be affected by individual student decisions, but more likely related to government expansion plans.

APTable 3.4 and APTable 3.5 show the estimated results using number of college teachers as a measure for the college expansion and its impact on migration at ages 17-20 and in the 3 years after graduation. We use the same models as those used in Table 3.3 and Table 3.4, and replace the variable local college admissions and the college admissions in other provinces by local number of college teachers and the number of college teachers in other provinces.

The estimated results in APTable 3.4 confirm our conclusions from Table 3.3 that the college expansion has a positive effect on migration at 17-20 for both college-bound students and non-college-bound students. In APTable 3.5, we find a significant negative effect on the interaction between college degree and local number of college teachers. This supports our conclusion in Table 3.4 that due to the competition effect, college graduates are more likely to migrate after graduation if there is a large supply increase in skilled labor in the province of graduation.

3.6.3 Specifying the Main Reason for Migration

To further separate the impact of college expansion from the impact of regional economic growth on migration, we re-define the outcome variable using the “migration reason” information. Specifically, for migration decisions made for those 17-20, we re-define the outcome dummy to equal to 1 only when the migration for ages 17-20 is for studying, and 0 otherwise. The regression results are listed in APTable 3.6.

As can be seen from the table, when restricting migration to moves taken for education, it does not change our basic conclusions from Table 3.3. It still suggests that an increase in college admissions in local provinces reduces the probability of inter-provincial migration for both college graduates and non-college graduates. Although the estimated coefficient for the interaction term between personal college degree attainment and college admissions is negative, suggesting that those who attend colleges are even less likely to move due to the increase in local admissions, it is not statistically significant. In contrast to what we observe in Table 3.3, the coefficient of the college dummy in this table suggests that if a person has obtained a college degree by the time of the survey, then he/she is more likely to migrate for education in the age range 17-20. Table 3.3 suggests the estimated coefficient

is negative, which is probably because most non-college graduates migrate for work in these ages, as we observe in Figure 3.4. Also, before adding the college dummy into the model, the coefficient for Hukou is significantly negative, which suggests that people with agricultural Hukou are less likely to migrate for education at these ages. After controlling the college dummy, the negative effect of having an agricultural Hukou disappears, indicating that having a non-agricultural Hukou increases the probability of migration for education because it reduces the chance of obtaining a college degree.

We also re-define the migration dummy after graduation to be 1 only when the migration is for work. By re-defining the migration dummy in this way, we may better understand the effect of an increase in the supply of skilled labor on labor-related migration decisions. The results are listed in APTable 3.7. The results are similar to those in Table 3.4, implying that the increase of local skilled labor supply drives college graduates to migrate to other provinces, but it induces non-college graduates not to migrate. We also find a significant positive effect of an increase of college admissions in other provinces on the migration decision of non-college graduates for work. These results further support the view that skilled workers and less-skilled workers are complementary to each other.

3.7 Conclusion

The college expansion in China not only provides more opportunities to students to attend college, but also affects people's migration. This chapter uses the migration history information for individuals from CLDS 2012 to study how the college expansion affects migration decisions. Firstly, an increase in outside college admissions increases the probability of inter-provincial migration for college-bound students at ages 17-20. We call this an "enrollment effect" since college-bound students migrate to attend college in other

provinces as the aggregate admissions increase. However, the “enrollment effect” of local admissions on migration for ages 17-20 is not clear from our analysis. The estimated results suggest that those who did not attend college are less likely to migrate as local admissions increase and little evidence suggests that the effect for college-bound students is significantly different. Since college expansion also brings economic opportunities to an area, it may induce non-college graduates to stay. We also find that as college admission increases, it increases the probability of migration for college graduates after graduation. We call this a “competition effect” since new college graduates face great competition from having more potential college graduates in the local market. The results suggest that when facing such competition, college graduates choose to migrate to other provinces. In addition, we use wild bootstrap test to re-estimate the linear models and obtain similar conclusions. To further exclude the effect of entering the WTO on migration, we treat those who migrated to eastern provinces as non-migrants, and it does not change our results. Using the number of college teachers as a proxy for the college expansion and restrict the reason for migration, we also obtain similar results. We use the conditional Logit model to estimate the likelihood of the migration area and find that an increase of college admissions in outside region induces people’s migration for ages 17-20.

Table 3.1 Summary Statistics on the CLDS 2012 Sample

	All	Birth Cohort 1952-1962	Birth Cohort 1963-1972	Birth Cohort 1973-1982
All Sample				
Age	37.57 (13.48)	54.83 (3.38)	44.35 (2.86)	34.73 (2.93)
Male	0.51 (0.50)	0.51 (0.50)	0.51 (0.50)	0.51 (0.50)
Schooling	9.14 (3.28)	8.02 (3.53)	8.91 (3.08)	9.49 (3.17)
Migrant Sample				
Migration Rate	11.79	5.28	10.03	20.03
Reason of First Migration				
Political	2.90	13.91	0.09	0.00
Family	9.44	15.45	9.78	8.18
Work	70.98	37.25	75.27	79.10
Education	4.65	0.96	3.91	6.05
Others	12.02	32.43	10.95	6.66
First Migration Age	21.43 (4.76)	21.05 (5.25)	21.83 (5.02)	21.25 (4.48)
First Migration Destination				
Eastern	62.70	58.99	58.51	66.50
Central	20.44	27.85	24.42	16.41
Western	16.86	13.16	17.06	17.09
Schooling	10.72 (3.20)	10.51 (3.28)	10.26 (3.28)	11.05 (3.08)

Note: This table lists the means and standard deviations for the CLDS sample in terms of demographic information (age, gender, years of schooling), migration rate, and reasons of migration by birth cohort. It also lists the means and standard deviations of demographic information and the distribution of migration region for migrant sample by birth cohort.

Table 3.2 Summary Statistics for Migrants by Migration Reason and by Migration Cohort

Migration Cohort	Family		Work		Education	
	1978-1998	1999-2012	1978-1998	1999-2012	1978-1998	1999-2012
Age at Migration	17.25 (7.06)	21.10 (4.64)	19.74 (3.91)	19.76 (3.87)	16.49 (3.89)	17.70 (3.68)
Schooling	9.91 (3.81)	11.19 (3.02)	10.16 (3.06)	10.83 (3.14)	13.77 (1.93)	13.64 (2.36)
Male	0.31 (0.47)	0.25 (0.43)	0.65 (0.48)	0.59 (0.49)	0.73 (0.45)	0.53 (0.50)
Hukou Status at Birth	0.90 (0.30)	0.87 (0.34)	0.92 (0.28)	0.92 (0.28)	0.65 (0.48)	0.76 (0.43)
Destination Region						
Eastern	51.84	57.98	68.65	75.22	40.85	42.42
Central	27.30	31.64	17.94	12.43	26.80	35.76
Western	20.86	10.38	13.41	12.35	32.36	21.82
Origin Region						
Eastern	53.98	60.62	34.70	36.52	42.09	39.36
Central	24.79	29.01	37.49	36.11	20.31	35.33
Western	21.23	10.37	27.81	27.37	37.60	25.31

Note: This table lists the means and standard deviations for demographic information (gender, age and years of schooling at migration, Hukou status at birth) and destination regions by migration cohort (1978-1998, 1999-2012).

Table 3.3 Estimated Results of Equation (3.6) on Migration Decision for Ages 17-20

Mig_1720	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
	(1)	(2)	(3)	(4)
Col*otherAD				0.110** (0.041)
Col*AD				0.007 (0.010)
AD	-0.019** (0.008)	-0.020* (0.011)	-0.019* (0.011)	-0.020* (0.010)
otherAD	0.042 (0.088)	0.019 (0.076)	0.013 (0.073)	0.008 (0.071)
Col			-0.036** (0.014)	0.010 (0.016)
Labor Mkt Var.	No	Yes	Yes	Yes
Interaction	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal Characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.054	0.056	0.058	0.061

Note: This table reports the estimation results of equation (3.6) of migration for ages 17-20. Col is the dummy for whether a person obtained a college degree when taking the survey; AD is the measure of the number of college admissions in the province the person was resided at age 16, and otherAD is the measure of the weighted average of the number of college admissions in provinces other than the person was resided at age 16. The weighting is based on equations (3.7) and (3.8) (see APTable 3.1 for more details). Control variables include labor market variables (SWAGE, UWAGE, UR), their interactions with Col, and population measured in provincial level in the year of aged 16, personal characteristics (gender, FS, Hukou), and provincial fixed effects and year fixed effects (the definition for each variable listed in the table is listed in APTable 3.1.). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in age 16, and were estimated using the wild bootstrap method⁴².

⁴² The wild bootstrap test is proposed in Cameron et al. (2008, 2010) to correct the bias of standard errors due to small number of clusters (Stata command “boottest”).

Table 3.4 Estimated Results of Equation (3.6) on Migration Decision Within 3 Years After Graduation

Mig_gra	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
	(1)	(3)	(3)	(4)
Col*otherAD				-0.027 (0.035)
Col*AD				0.041* (0.021)
AD	-0.028 (0.017)	-0.022 (0.018)	-0.022 (0.018)	-0.032* (0.019)
otherAD	0.066** (0.026)	0.060** (0.027)	0.058** (0.029)	0.033 (0.037)
Col			-0.010 (0.030)	0.072 (0.068)
Labor Mkt Var.	No	Yes	Yes	Yes
Interaction	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal Characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.067	0.087	0.088	0.090

Note: This table reports the estimation results of equation (3.6) for migration after graduation. Col is the dummy for whether a person obtained a college degree when taking the survey; AD is the measure of the number of college admissions in the aged 16 of the province a person i was graduated from, and otherAD is the measure of the weighted average of the number of college admissions in the aged 16 of provinces other than the person graduated from. The weighting is based on equations (3.7) and (3.8) (see APTable 3.1 for more details). Control variables include labor market variables (SWAGE, UWAGE, UR), their interactions with Col, and population measured in provincial level in the year of graduation, personal characteristics (gender, FS, Hukou), and provincial fixed effects and year fixed effects (the definition for each variable listed in the table is listed in APTable 3.1.). *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in the year of graduation, and were estimated using the wild bootstrap method.

Table 3.5 Marginal Effects on Migration Decision on Migration Decision For Ages 17-20 Using McFadden's Conditional Logit Model

Marginal Effect (Mig in 17-20)	(1)	(2)	(3)
Col *AD*(1-Home)			0.01
Col*AD		0.19	0.10
AD*(1-Home)			0.02
AD	0.02	0.01	0.01
Distance	-0.12*	-0.11*	-0.07**
Home	0.68***	0.68***	0.11***
Central			
Gender	-0.02	-0.03	-0.03
FS	0.01	0.01	0.01
Hukou	-0.13**	-0.18**	-0.17**
Col	0.17***	0.12***	0.10**
Western			
Gender	0.02	0.01	0.01
FS	-0.01	-0.01	0.00
Hukou	-0.07*	-0.03*	-0.02*
Col	0.13***	0.02	0.01
Area FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	4,270	4,270	4,270
Log L	-871.2	-869.4	-866.0

Note: This table reports the marginal effects of equation (3.10) for ages 17-20 using the McFadden's conditional Logit model. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. AD stands for the average annual college admissions for provinces in a region at the year of age 16. Home denotes if the region is the residential region at age 16. Distance is the measure of average distance between the provinces of the original region and the provinces of a particular destination region (unit: 1,000 km) ($Dist_{i,j} = \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} d_{k,l} / (N_i * N_j)$), where k is the origin province within region i , and l is a destination province within region j). Control variables include the average wage of skilled-labor-intensive industries in the home region in the calendar year when the person is age 16; the average wage of unskilled-labor-intensive industries in the home region in the calendar year when the person is age 16; the unemployment rate in the home region in the calendar year when the person is age 16; the total population in the home region in the calendar year when the person is age 16; gender, year's of father's schooling, Hukou at birth, and college degree attainment. Each of the macro-level variables is coded as the differences between raw values and the overall mean.

Table 3.6 Marginal Effects on Migration Decision Within 3 Years After Graduation Using McFadden's Conditional Logit Model

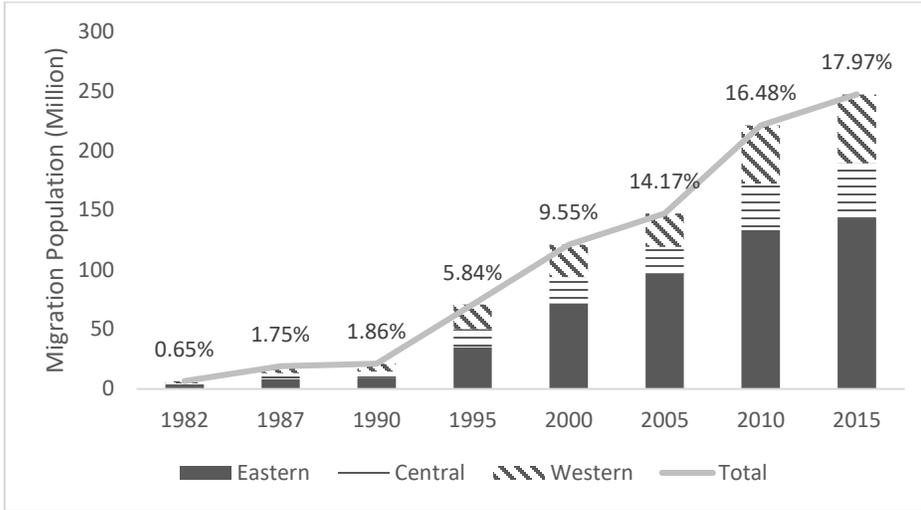
Marginal Effect (Mig after graduation)	(1)	(2)	(3)
Col *(1-Home)			0.01
Col*AD		-0.09*	-0.07*
AD*(1-Home)			0.03
AD	0.05	0.03	0.01
Distance	-0.20**	-0.15**	-0.10**
Home	0.81***	0.73***	0.68***
Central			
Gender	0.01	0.04	0.01
FS	0.01	0.01	0.01
Hukou	0.12**	0.08**	0.06
Col	-0.10*	-0.04**	-0.03
Western			
Gender	0.01	0.01	0.01
FS	0.01	0.01	0.01
Hukou	0.06**	0.02**	0.01
Col	-0.09***	-0.09***	-0.02
Area FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	4,270	4,270	4,270
Log L	-446.1	-441.9	-420.0

Note: This table reports the marginal effects of equation (3.10) for migration in 3 years after graduation using the McFadden's conditional Logit model. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. AD stands for the average annual college admissions for provinces in a region at the year of age 16. Home denotes if the region is the residential region of graduation. Distance is the measure of average distance between the provinces of the original region and the provinces of a particular destination region (unit: 1,000 km) ($Dist_{i,j} = \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} d_{k,l} / (N_i * N_j)$), where k is the origin province within region i , and l is a destination province within region j). Control variables include the average wage of skilled-labor-intensive industries in the home region in the calendar year when the person graduated; the average wage of unskilled-labor-intensive industries in the home region in the calendar year when the person graduated; the unemployment rate in the home region in the calendar year when the person graduated; the total population in the home region in the calendar year when the person graduated; gender, year's of father's schooling, Hukou at birth, and college degree attainment. Each of the macro-level variables is coded as the differences between raw values and the overall mean.

Table 3.7 Calculated Chi-square for Models with Full Sample and Subsample in Table 3.5 and Table 3.6

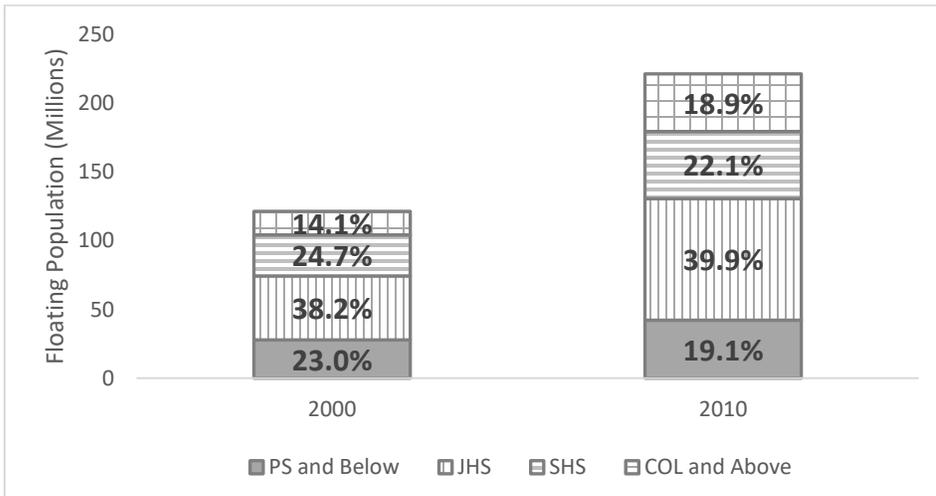
For migration for ages 17-20	Drop eastern region	Drop central region	Drop western region	Chi-square at d.f. 17 (p=0.05)
χ^2	6.27	12.76	24.47	27.59
	6.41	19.25	16.95	
	1.87	5.18	14.90	
For migration after graduation	Drop eastern region	Drop central region	Drop western region	
χ^2	1.26	26.34	0.59	
	1.56	11.09	0.16	
	2.03	5.83	1.81	

Figure 3.1 The Floating Population and Its Regional Distribution in China



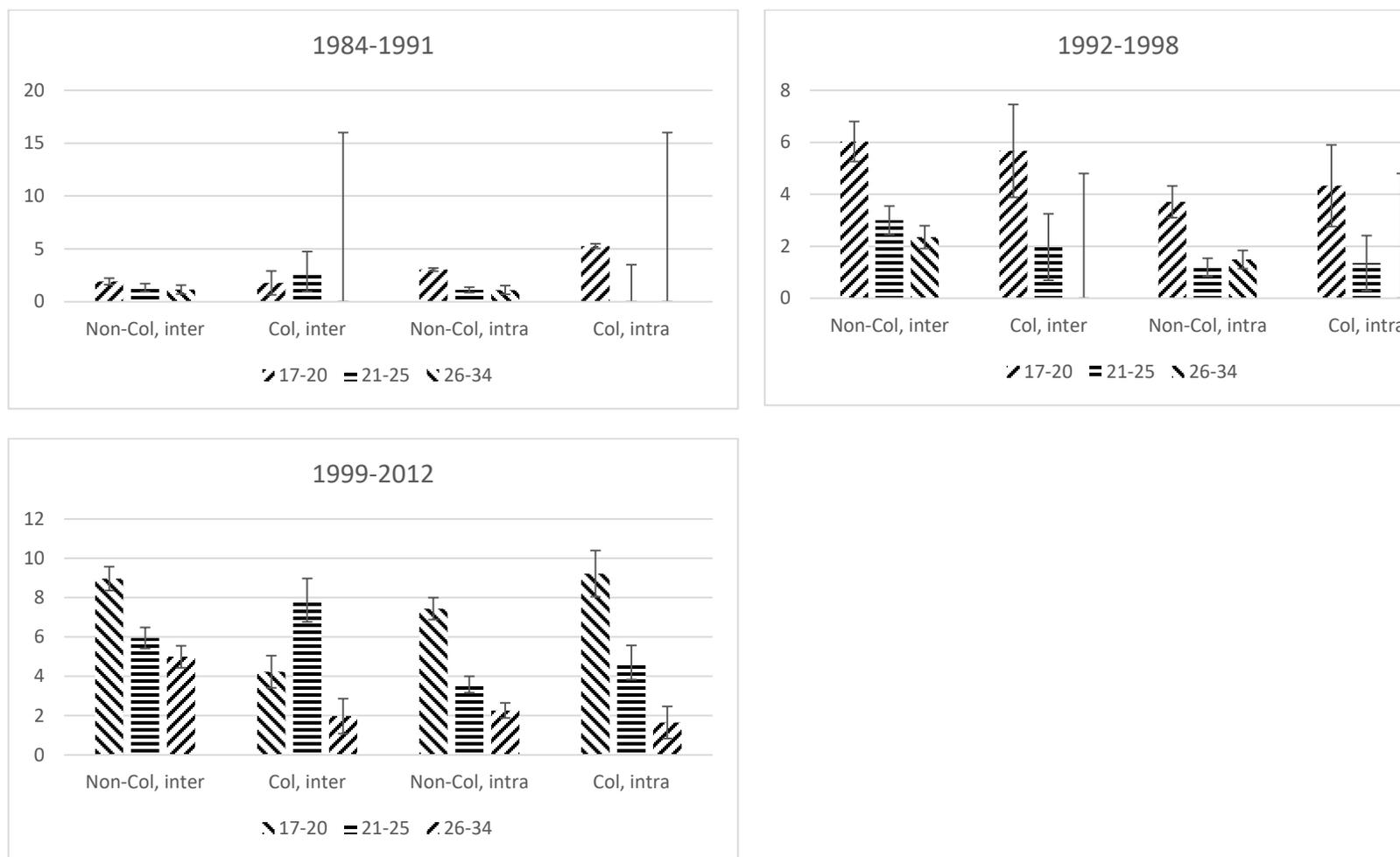
Note: This chart shows the total size and the regional distribution of the floating population in China from 1982 to 2015. The y-axis denotes the population (unit: 1 million), the solid area in each bar denotes the proportion where the migration destination is in the eastern provinces, the patterned area with horizontal lines denotes central provinces, and the patterned area with diagonal slashes denotes western provinces. The trend line plots the floating population, and the percentage above each bar denotes the percentage of migrants in total population of the corresponding year. (Source: China's Census in 1982, 1990, 2000, 2010; and 1% National Sample Survey in 1987, 1992, 2005 and 2015).

Figure 3.2 Educational Distribution of Migrants in China



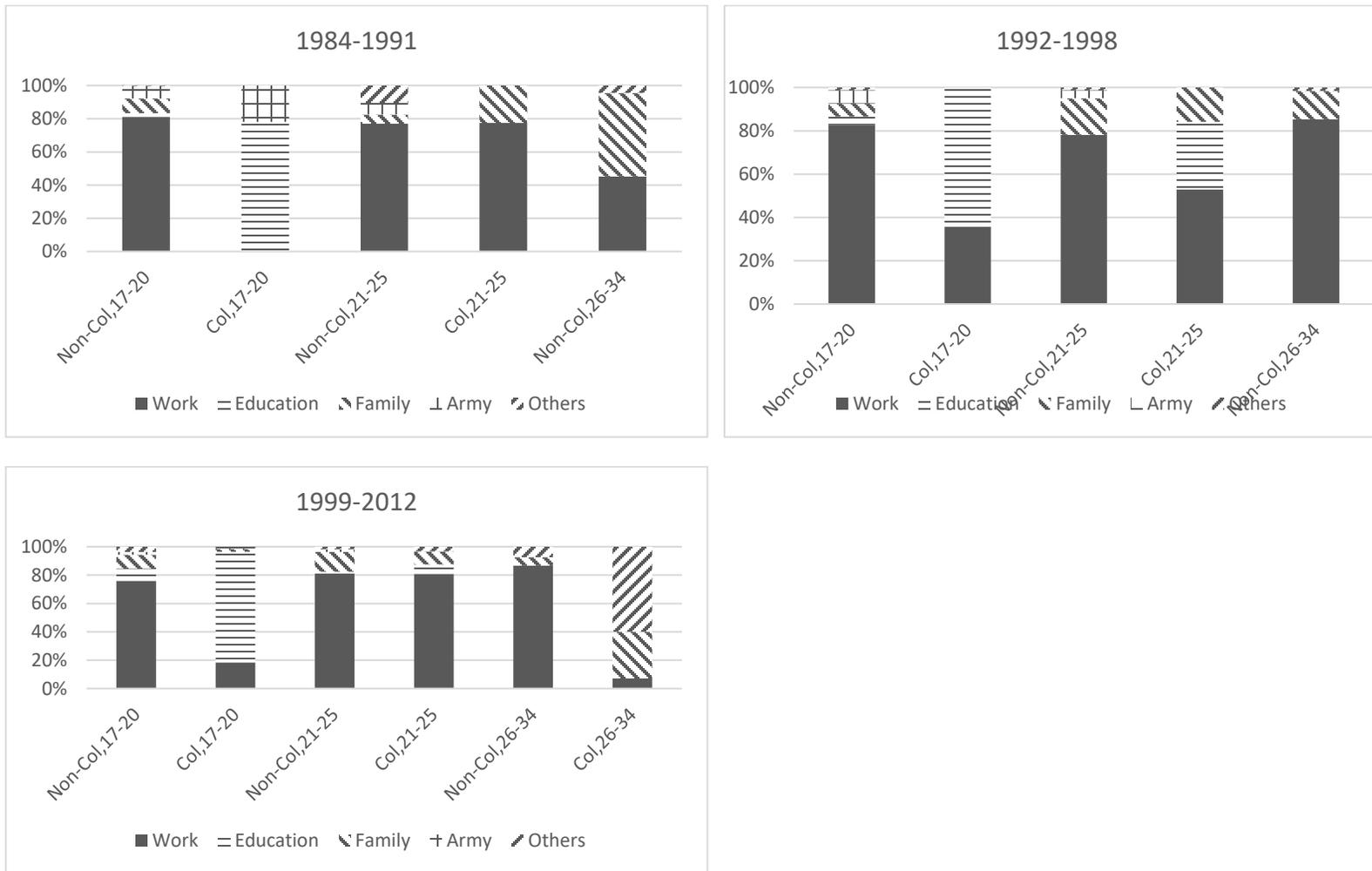
Note: This chart shows the total size and the educational distribution of the floating population in China in 2000 and 2010. The y-axis denotes the size of the floating population (million), grey area in each bar denotes the proportion of this population with primary school or less, the vertical line pattern denotes the proportion of having middle school education, horizontal line area denotes high school education, and cross hatch area denotes college education and above (Source: China's Census in 2000 and 2010).

Figure 3.3 Inter-provincial and Intra-provincial Migration Rates by Age Group and by Education in 3 Time Periods



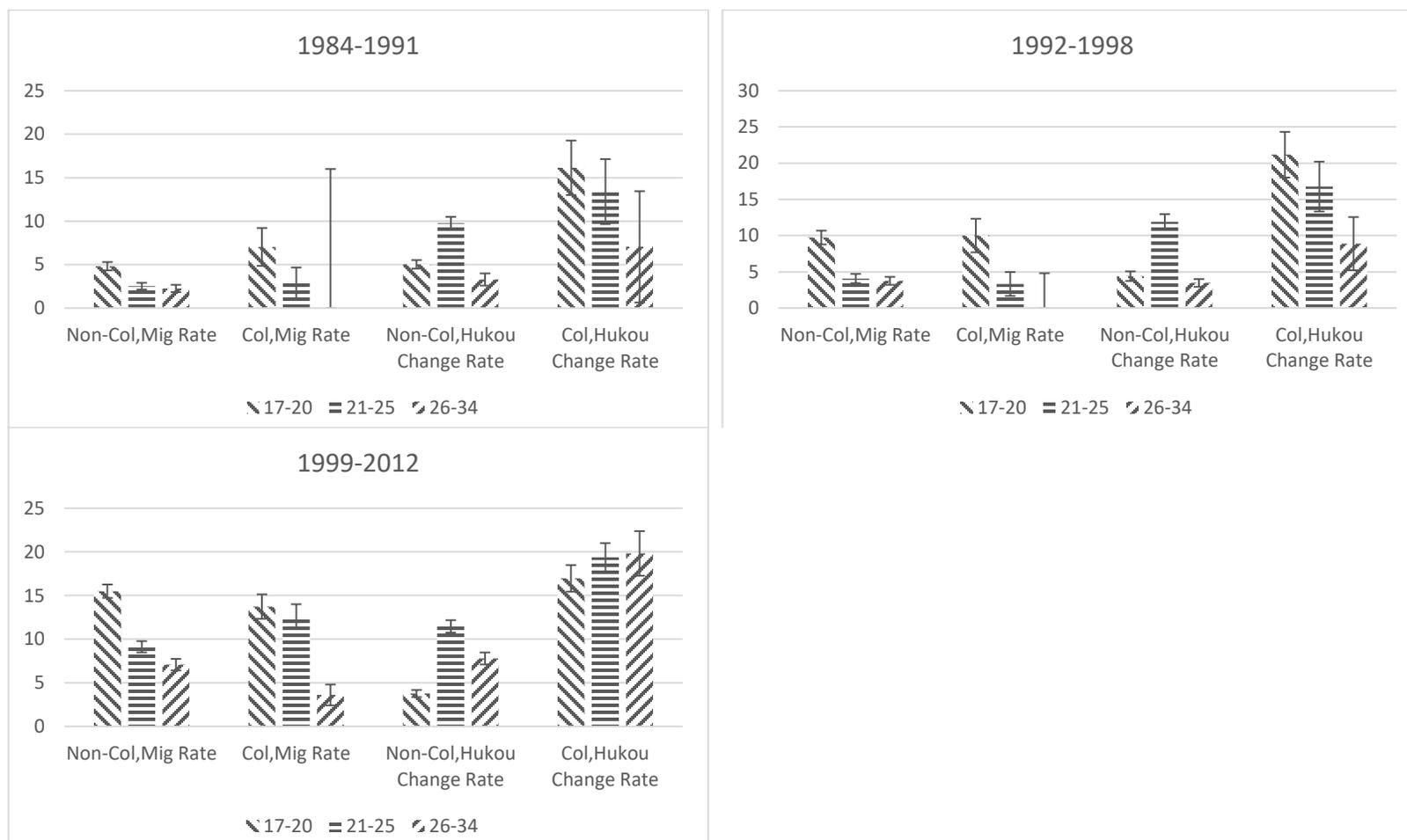
Note: This chart shows the inter-province and intra-province migration rates and their 95% confidence intervals by education and by age group for migration in years 1984-1991, 1992-1998, and 1999-2012. The y-axis denotes the migration rate in percent, backward slash bars denote the migration rates for ages 17-20, horizontal line bars denote the migration rates for ages 21-25, and the forward slash bars denote the migration rates for ages 26-34. The migration rates are calculated using the number of people who migrated in a time period and migrated in a specified age range and who were the specified ages entirely within the time period, divided by the number of respondents who were the specified ages entirely within the time period. The migration rates are calculated using the CLDS 2012 sample.

Figure 3.4 Distribution of Migration Reasons by Education and Age Group



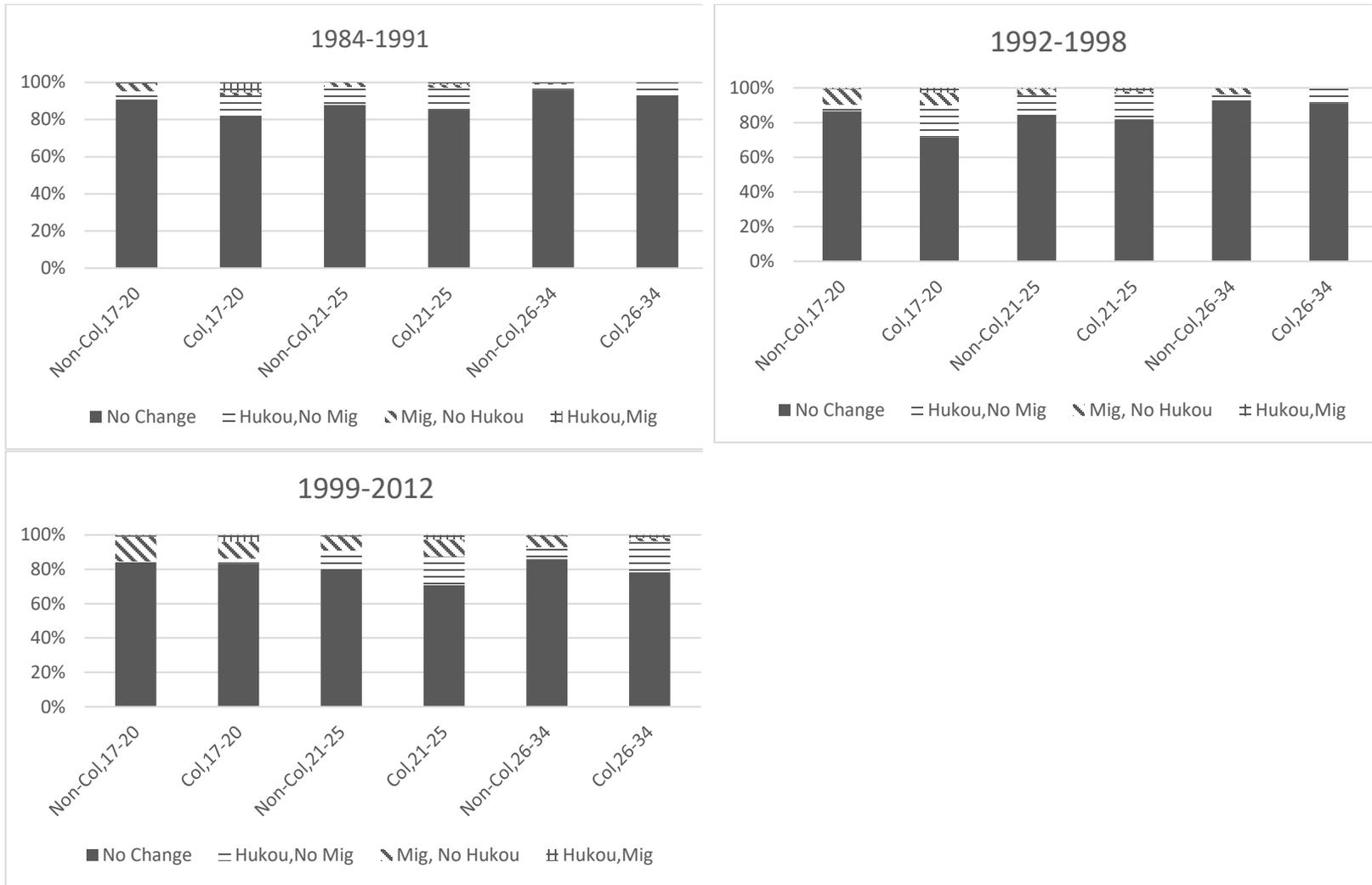
Note: This chart shows the distribution of migration reasons by education and by age group in 1984-1991, 1992-1998, and 1999-2012. The y-axis denotes the proportion of each migration reason among all migrants in an age group. From bottom to the top of each bar, the migration reasons can be summarized as for work, for education, for family, for join army, and other reasons. We use the migration history data in the CLDS 2012 to calculate these proportions.

Figure 3.5 Migration Rates and Hukou Change Rates by Age Group and by Education



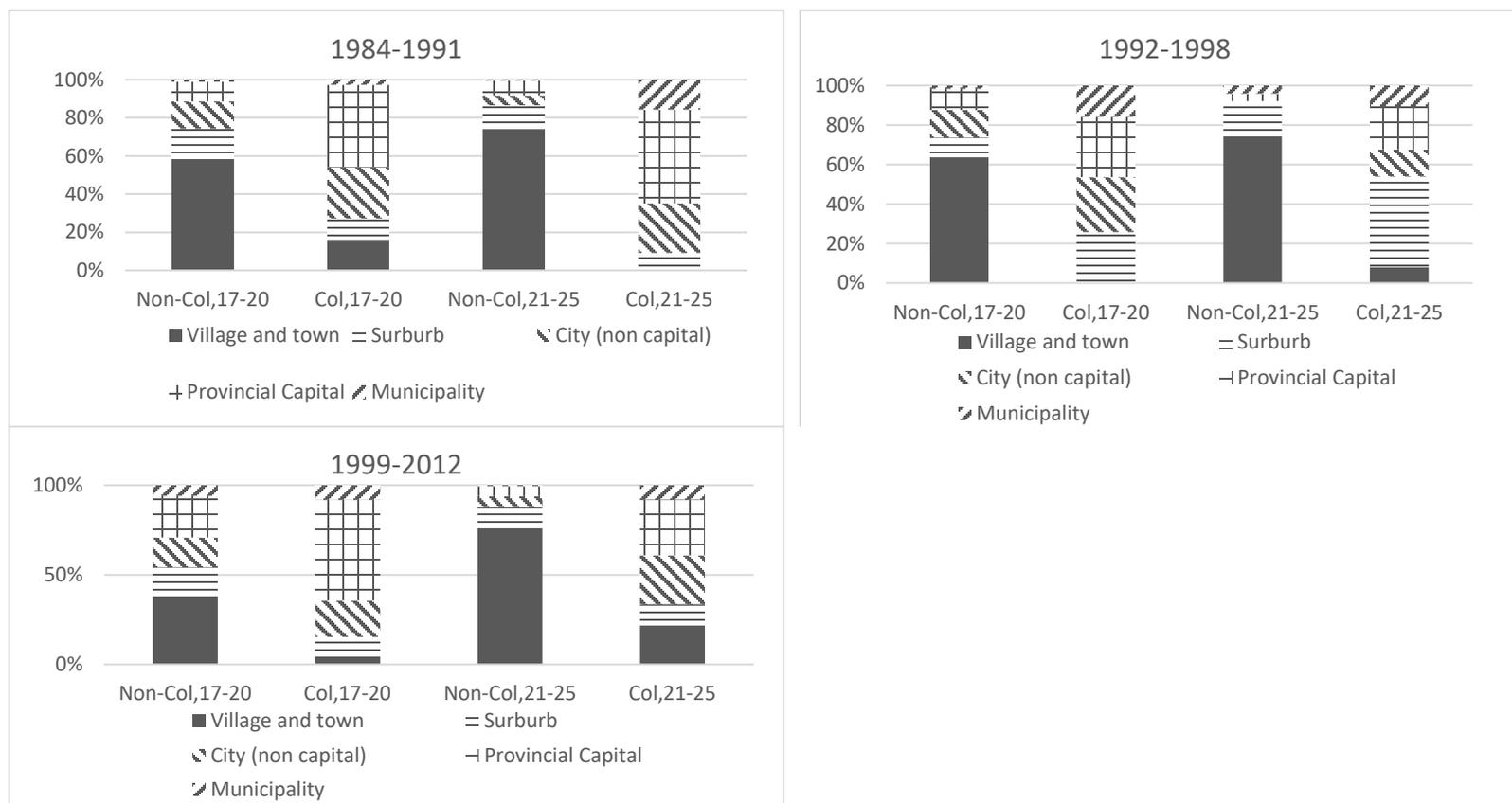
Note: This chart shows the migration rates and Hukou change rates by education and by age group in 1984-1991, 1992-1998, 1999-2012. The y-axis denotes migration rate in percentage, backward slash bars denote the rates for ages 17-20, horizontal line bars denote the rates for ages 21-25, and the forward slash bars denote the rates for ages 26-34. The migration rates are calculated using the number of people in a specified age range, who migrated in a given time period and spent the entire specified ages in the time period, divided by the number of respondents whose specified age range is covered by the time period. The Hukou change rates are calculated using the number of people who changed Hukou in a given time period and in a specified age range, divided by the number of respondents in the specified age range and time period. The migration rates and Hukou change are calculated using the CLDS 2012 sample.

Figure 3.6 Proportion of Different Situations of Migration and Hukou Change by Age Group and by Education



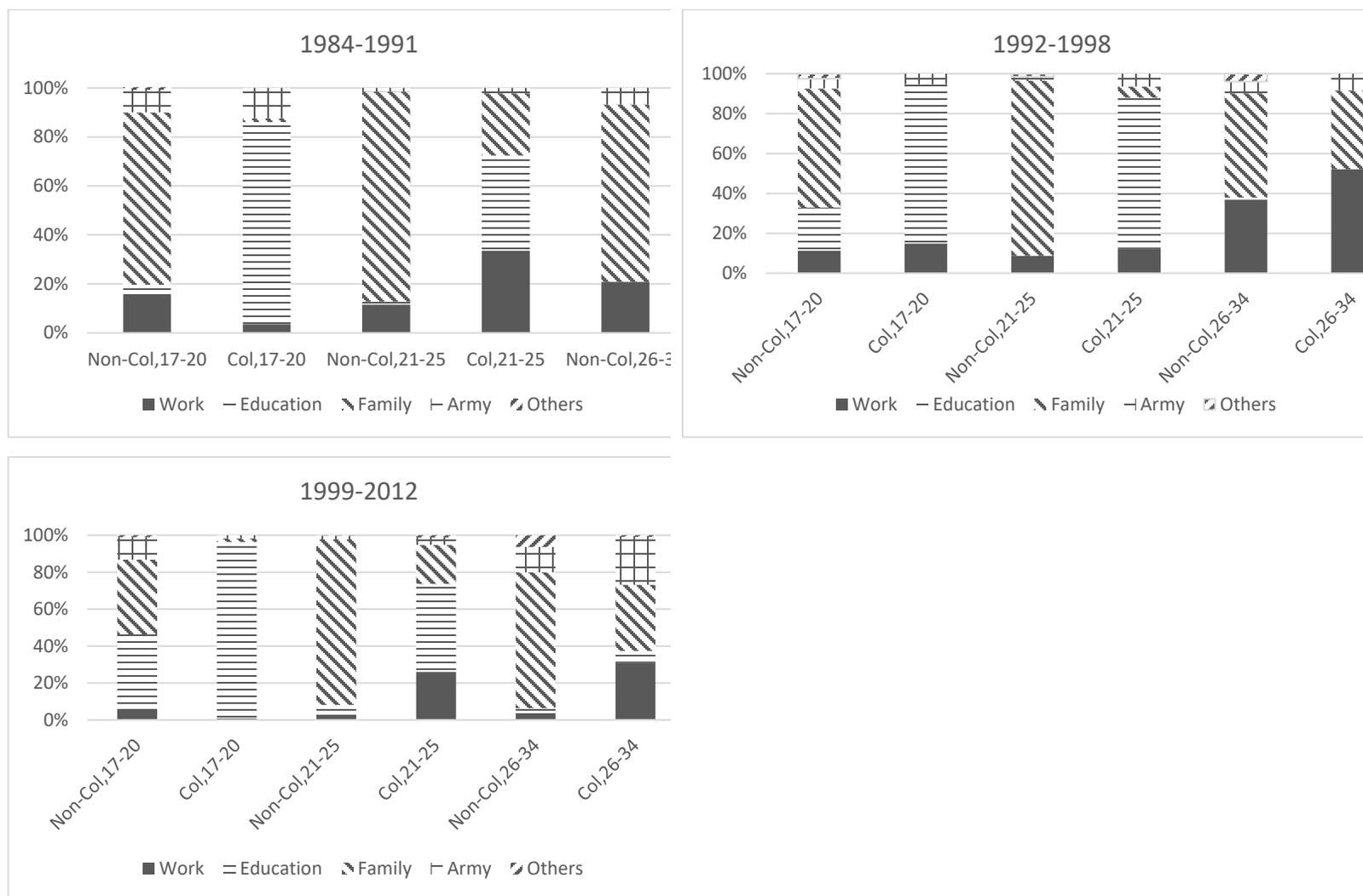
Note: This chart shows the distribution of migration and Hukou change by education and by age group in 1984-1991, 1992-1998, 1999-2012. The y-axis denotes proportion of each migration reason among all migrants in an age group. From bottom to the top of each bar, the solid area denotes the percentage of people who did not migrate or change Hukou in each age range, the backslash area denotes the percentage of people who changed Hukou but did not migrate, the checked area denotes the percentage of people who migrated but did not change Hukou, and the forward slash area denotes the percentage of people who both migrated and changed Hukou.

Figure 3.7 Distribution of Regional Type of Hukou Change by Destinations, Age Group, and Education



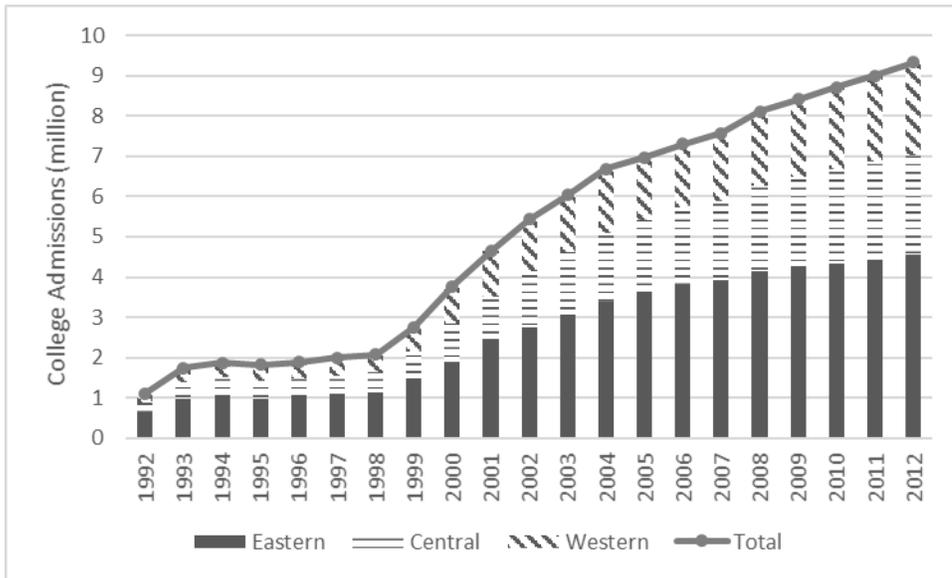
Note: This chart shows the distribution of region type of Hukou destination by education and by age group in 1984-1991, 1992-1998, and 1999-2012. We use the Hukou change history data in the CLDS 2012 to calculate these proportions. The y-axis denotes proportion of each destination region type among all Hukou changes in an age group. The solid area denotes destinations in small villages and towns. In terms of China's administrative definitions, it includes villages and townships that are not the administrative center regions of counties. The area with horizontal lines denotes suburban areas around cities, and in term of administrative definitions, it includes townships that are the administrative center regions of counties but are not in the administrative centers of prefectural-level cities. The area with backslash denotes cities, excluding province capitals. In terms of the administrative terminology, this category corresponds to the counties that are administrative centers of prefectural-level cities and are not administrative centers of regions in the province. The area with check pattern denotes provincial capitals. In administrative terms, it denotes administrative center cities of provinces. The area with the forward slash denotes municipalities, which identifies the four large cities (Beijing, Shanghai, Tianjin, and Chongqing) that are administratively treated as provinces. The administrative division in China includes five levels (highest to lowest): province level (including the four municipalities), prefectural level, county level, township level, and village.

Figure 3.8 Distribution of Hukou Change Reason by Age Group and by Education



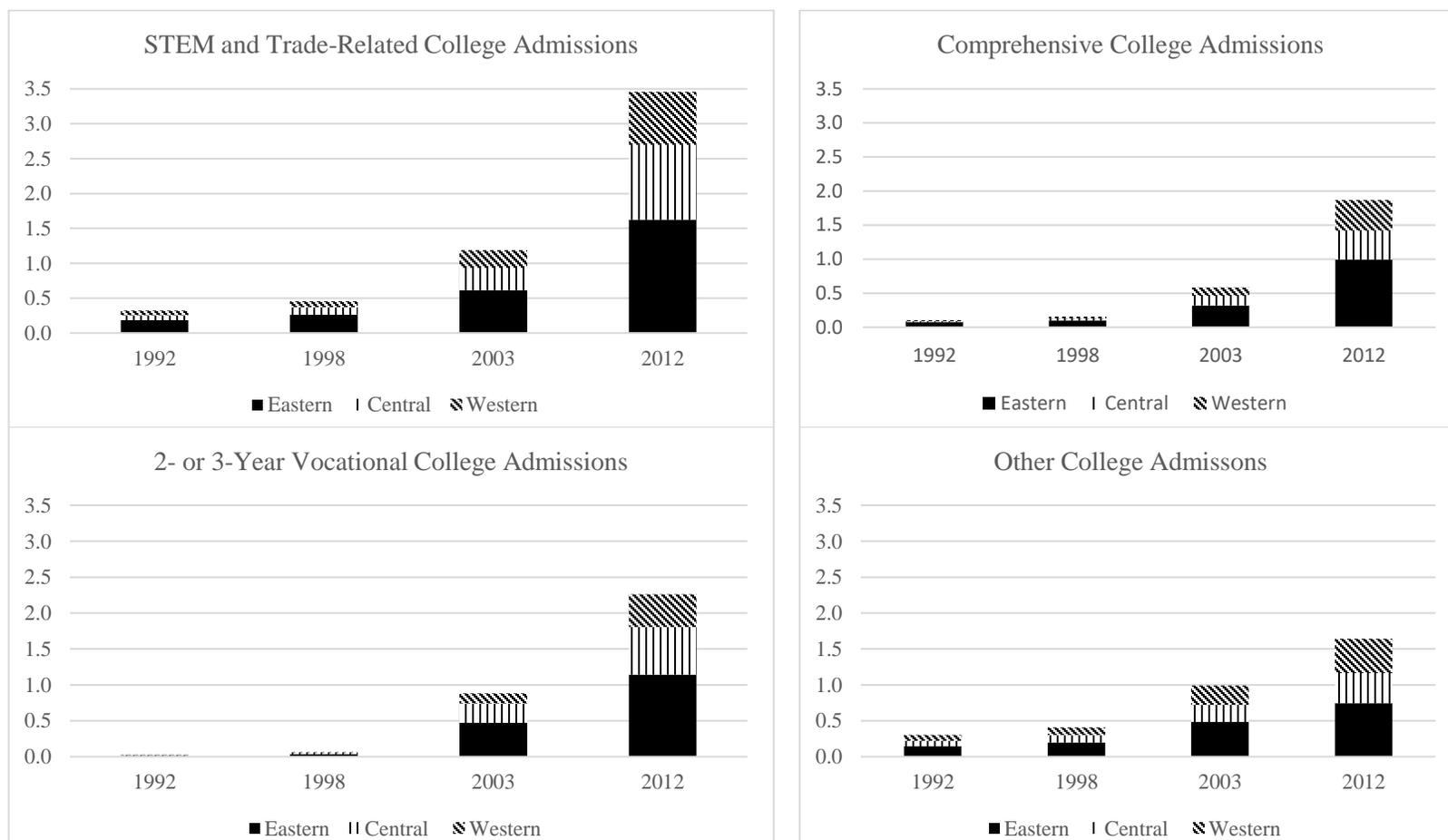
Note: This chart shows distribution of Hukou change reasons by education and by age group in 1984-1991, 1992-1998, 1999-2012. The y-axis denotes proportion of each Hukou change reason among all people who changed Hukou in an age group. From bottom to the top of each bar, the reasons can be summarized as for work, for receiving education, for family, for political reasons, and other reasons. We use the Hukou change history data in the CLDS 2012 to calculate these proportions.

Figure 3.9 Total Number and Regional Distribution of College Admissions by Year



Note: This chart shows the total number and the regional distribution of college admissions in China from 1992 to 2012. The y-axis denotes the number of college admissions (unit: 1 million) The solid black area in each bar denotes the proportion of college admissions in the eastern provinces, the hyphen area denotes the proportion in the central provinces, and the area with diagonal slashes denotes the proportion in the western provinces (Source: China Education Yearbook from 1996 to 2015).

Figure 3.10 Number of College Admissions by Region and Type in China



Note: This chart shows the total number and the regional distribution of college admissions in China from 1992 to 2012 by college type (from upper left to lower right: STEM and Trade-related colleges, comprehensive colleges, 2- or 3-year vocational colleges, and other colleges). The y-axis denotes the number of college admissions (unit: 1 million) The solid black area in each bar denotes the proportion of college admissions in the eastern provinces, the hyphen area denotes the proportion in the central provinces, and the area with diagonal slashes denotes the proportion in the western provinces (Source: Educational Statistics Yearbook of China, 1992, 1998, 2003, 2015).

3.8 Appendix

APTable 3.1 Variable Description

	Mean (SD)	Min	Max	Description	Source
AD	1.65 (1.46)	0.01	6.91	Number of college admissions in residential province at age 16 (for 17-20 migration) (unit: 100,000)	China Education Yearbook 1992-2013
SWAGE	10.31 (8.08)	2.17	50.79	Average annual wage for the 5 most skilled labor-intensive industries in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (unit: 1,000 Yuan)	China Labor Statistical Yearbook 1992-2013
UWAGE	8.26 (5.74)	2.04	3.38	Average annual wage for the 5 most unskilled labor-intensive industries in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (unit: 1,000 Yuan)	China Labor Statistical Yearbook 1992-2013
UR	3.41 (0.98)	0.30	7.40	Unemployment rate in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (%)	China Labor Statistical Yearbook 1992 to 2013
POP	4.09 (2.62)	0.23	10.64	Population in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (unit: 10 million)	China Statistical Yearbook 1992 to 2013
OtherAD	1.65 (1.15)	0.08	4.83	Weighted average number of college admissions in province other than the province at age 16 (unit: 100,000)	See equations (3.7)-(3.9)

OtherSWAGE	10.31 (7.88)	1.02	40.00	Weighted average annual wage for the 5 most skilled labor-intensive industries in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (unit: 1,000 Yuan)	See equations (3.7)-(3.9)
OtherUwage	8.26 (5.99)	1.00	30.58	Weighted average annual wage for the 5 most unskilled labor-intensive industries in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (unit: 1,000 Yuan)	See equations (3.7)-(3.9)
OtherUR	3.41 (1.06)	0.92	5.82	Weighted average unemployment rate in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (%)	See equations (3.7)-(3.9)
OtherPOP	4.09 (1.51)	0.91	6.87	Weighted average population in province at age 16 (for “17-20 migration analyses”) or province in the year of graduation (for “migration after graduation analyses”) (unit: 10 million)	See equations (3.7)-(3.9)
Male	0.55 (0.50)	0	1	A dummy for gender (male=1, female=0)	CLDS 2012 (age>=20)
Age (>=20)	39.82 (12.31)	20	64	Age at the time of the survey (2012)	CLDS sample
Col	0.11 (0.31)	0	1	Dummy for obtaining a college degree (yes=1, no=0)	CLDS 2012 (age>=20)
Hukou	0.84 (0.37)	0	1	Dummy for Hukou type at birth (agricultural=1, non-agricultural=0)	CLDS 2012 (age>=20)
Fschool	6.12 (4.30)	0	19	Years of father’s schooling	CLDS 2012 (age>=20)

APTable 3.2 Regression Results for Non-Eastern Migration for Ages 17-20

Mig_1720	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
	(1)	(2)	(3)	(4)
Col *otherAD				0.078 (0.056)
Col*AD				0.003 (0.012)
otherAD	-0.010 (0.051)	-0.004 (0.053)	-0.003 (0.054)	-0.014 (0.058)
AD	-0.014* (0.007)	-0.010* (0.005)	-0.010* (0.006)	-0.010* (0.006)
Col			0.019 (0.013)	0.028 (0.020)
Labor Mkt Var.	No	Yes	Yes	Yes
Interactions	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.043	0.045	0.047	0.047

Note: This table reports the estimation results of equation (3.6) of migration for ages 17-20, excluding migration to the eastern provinces. The definition for each variable listed in the table is the same as those shown in Table 3.3. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in age 16, and were estimated using the wild bootstrap method.

APTable 3.3 Regression Results for Non-Eastern Migration in 3 Years After Graduation

Mig_gra	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
	(1)	(2)	(3)	(4)
Col *otherAD				0.038 (0.026)
Col*AD				0.020* (0.011)
otherAD	0.006 (0.016)	-0.001 (0.014)	-0.002 (0.015)	-0.001 (0.001)
AD	-0.003 (0.011)	0.002 (0.011)	0.002 (0.011)	-0.005 (0.011)
Col			-0.006 (0.019)	0.030 (0.020)
Labor Mkt Var.	No	Yes	Yes	Yes
Interactions	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.056	0.067	0.075	0.079

Note: This table reports the estimation results of equation (3.6) for migration after graduation, excluding migration to the eastern provinces. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The definition for each variable listed in the table is the same as those shown in Table 3.3. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in the year of graduation, and were estimated using the wild bootstrap method.

APTable 3.4 Regression Results for Migration for Ages 17-20 Using the Number of Teachers as the Measure for College Expansion

Mig_1720	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
	(1)	(2)	(3)	(4)
Col *otherNT				0.408 (0.259)
Col*NT				-0.027 (0.090)
otherNT	-0.098 (0.047)	0.012 (0.041)	0.009 (0.040)	-0.227 (0.403)
NT	-0.117** (0.057)	-0.126* (0.077)	-0.126* (0.077)	-0.121* (0.073)
Col			-0.036** (0.014)	-0.008 (0.017)
Labor Mkt Var.	No	Yes	Yes	Yes
Interactions	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.053	0.054	0.057	0.059

Note: This table reports the estimation results of equation (3.6) of migration for ages 17-20. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. NT stands for number of college teachers in a person's residential province year at age 16 (origin province). The definition for each variable listed in the table is the same as those shown in Table 3.3. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in age 16, and were estimated using the wild bootstrap method.

APTable 3.5 Regression Results for Migration in 3 Years After Graduation Using the Number of Teachers as the Measure for College Expansion

Mig_gra	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Number of Teachers	(1)	(2)	(3)	(4)
Col *otherNT				0.035 (0.028)
Col*NT				0.068* (0.040)
otherNT	0.041** (0.019)	0.036** (0.019)	0.025 (0.020)	0.013 (0.022)
NT	-0.018 (0.012)	-0.014 (0.013)	-0.014 (0.013)	-0.015 (0.013)
Col			-0.040** (0.020)	0.010 (0.050)
Labor Mkt Var.	No	Yes	Yes	Yes
Interactions	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.063	0.065	0.065	0.067

Note: This table reports the estimation results of equation (3.6) for migration after graduation. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. NT stands for number of college teachers in a person's residential province year at age 16 (origin province). The definition for each variable listed in the table is the same as those shown in Table 3.3. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in the year of graduation, and were estimated using the wild bootstrap method.

APTable 3.6 Regression Results for Migration for Education Between Age 17 to 20

Mig_edu	Coef. (SE)	Coef. (SE)
	(1)	(2)
OtherAD	0.007** (0.003)	0.021*** (0.005)
AD	-0.004** (0.002)	-0.003* (0.002)
Labor Mkt Var.	No	Yes
Population	Yes	Yes
Personal characteristics	Yes	Yes
Provincial FE	Yes	Yes
Year FE	Yes	Yes
N	4,270	4,270
R ²	0.022	0.036

Note: This table reports the estimation results of equation (3.6) of migration for education between age 17 and 20. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable, Mig_edu, is defined as whether a person migrated to another province for ages 17-20, and the main reason of the migration is education. The definition for each of the other variables listed in the table is the same as those shown in Table 3.3. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in age 16, and were estimated using the wild bootstrap method.

APTable 3.7 Regression Results for Migration for Working in 3 Years After Graduation

Mig_work	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
	(1)	(2)	(3)	(4)
Col *otherAD				-0.015 (0.038)
Col*AD				0.036* (0.019)
otherAD	0.068*** (0.022)	0.067*** (0.023)	0.059** (0.024)	0.062*** (0.026)
AD	-0.031** (0.015)	-0.029** (0.016)	-0.030** (0.016)	-0.032** (0.017)
Col		-0.030 (0.020)	0.030 (0.020)	0.020 (0.019)
Labor Mkt Var.	No	No	Yes	Yes
Interactions	No	No	No	Yes
Population	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Provincial FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,270	4,270	4,270	4,270
R ²	0.089	0.097	0.091	0.093

Note: This table reports the estimation results of equation (3.6) for migration for work after graduation. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. The dependent variable, Mig_work, is defined as whether a person migrated to another province in three year after graduation, and the main reason of the migration is work. The definition for each of the other variables listed in the table is the same as those shown in Table 3.3. *** significance level <0.001; ** significance level <0.005; * significance level <0.01. Each of the macro-level variables is coded as the differences between raw values and the overall mean. The standard errors are clustered by the residential province in the year of graduation, and were estimated using the wild bootstrap method.

3.9 Reference

- Abbott, W. F., & Schmid, C. F. (1975). University Prestige and First-time Undergraduate Migration in the United States. *Sociology of Education*, 48(2), 168-185.
- Anjomani, A. (2002). Regional Growth and Interstate Migration. *Socio-Economic Planning Sciences*, 36(4), 239-265.
- Andrienko, Y., & Guriev, S. (2004). Determinants of Interregional Mobility in Russia. *Economics of Transition*, 12(1), 1-27.
- Bauer, T., & Zimmermann, K. (1995). *Integrating the East: The Labor Market Effects of Immigration* (Working Paper No. 5). Retrieved from American Institute for Contemporary German Studies.
- Bauer, T., & Zimmermann, K. F. (1997). Network Migration of Ethnic Germans. *International Migration Review*, 31(1), 143-149.
- Black, D., Gates, G., Sanders, S., & Taylor, L. (2000). Demographics of the Gay and Lesbian Population in the United States: Evidence from Available Systematic Data Sources. *Demography*, 37(2), 139-154.
- Böckerman, P., & Haapanen, M. (2013). The Effect of Polytechnic Reform on Migration. *Journal of Population Economics*, 26(2), 593-617.
- Borjas, G. J., & Bronars, S. G. (1991). Immigration and the Family. *Journal of Labor Economics*, 9(2), 123-148.

- Boyle, P. J., & Halfacree, K. H. (1995). Service Class Migration in England and Wales, 1980–1981: Identifying Gender-specific Mobility Patterns. *Regional Studies*, 29(1), 43-57.
- Cadwallader, M. T. (1992). *Migration and Residential Mobility: Macro and Micro Approaches*. University of Wisconsin Press.
- Carrington, W. J., Detragiache, E., & Vishwanath, T. (1996). Migration with Endogenous Moving Costs. *American Economic Review*, 86(4), 909-930.
- Champion, T., Fotheringham, S., Rees, P., Boyle, P., & Stillwell, J. (1998). *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence*. The Department of Geography, University of Newcastle upon Tyne.
- Chen, Y., & Rosenthal, S. S. (2008). Local Amenities and Life-cycle Migration: Do People Move for Jobs or Fun? *Journal of Urban Economics*, 64(3), 519-537.
- Cobb Clark, D. A. (1990). *Immigrant Selectivity: The Roles of Household Structure and US Immigration Policy* (Doctoral dissertation). Retrieved from <https://deepblue.lib.umich.edu/>
- Cooke, T. J., & Boyle, P. (2011). The Migration of High School Graduates to College. *Educational Evaluation and Policy Analysis*, 33(2), 202-213.
- Cushing, B., & Poot, J. (2004). Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling. *Papers in Regional Science*, 83(1), 317-338.
- DaVanzo, J. (1978). Does Unemployment Affect Migration? Evidence from Micro Data. *The Review of Economics and Statistics*, 60(4), 504-514.

Daveri, F., & Faini, R. (1999). Where Do Migrants Go? *Oxford Economic Papers*, 51(4), 595-622.

Ederveen, S., & Bardsley, N. (2003, Jan.24). *The Influence of Wage And Unemployment Differentials on Labour Mobility in the EU: A Meta-Analysis*. Ageing and Welfare Systems: What Have We Learned 2003 Conference, Brussels, Belgium.

Gottlieb, P. D., & Joseph, G. (2006). College-to-work Migration of Technology Graduates and Holders of Doctorates within the United States. *Journal of Regional Science*, 46(4), 627-659.

Graves, P. E., & Linneman, P. (1979). Household Migration: Theoretical and Empirical Result. *Journal of urban Economics*, 6(3), 383-404.

Greenwood, M. J. (1975). Research on Internal Migration in the United States: A Survey. *Journal of Economic Literature*, 397-433.

Greenwood, M. J. (1985). Human Migration: Theory, Models, and Empirical Studies. *Journal of Regional Science*, 25(4), 521-544.

Greenwood, M. J. (1997). Internal Migration in Developed Countries. *Handbook of Population and Family Economics*, 1, 647-720.

Faggian, A., McCann, P., & Sheppard, S. (2007). Human Capital, Higher Education and Graduate Migration: An Analysis of Scottish and Welsh students. *Urban Studies*, 44(13), 2511-2528.

Faggian, A., & Franklin, R. S. (2014). Human Capital Redistribution in the USA: the Migration of the College-bound. *Spatial Economic Analysis*, 9(4), 376-395.

- Fenske, R. H. (1972). *College Student Migration* (ACT Research Report No.54) Retrieved from the Research and Development Division, The American College Testing Program http://www.act.org/content/dam/act/unsecured/documents/ACT_RR54.pdf
- Fu, Y., & Gabriel, S. A. (2012). Labor Migration, Human Capital Agglomeration and Regional Development in China. *Regional Science and Urban Economics*, 42(3), 473-484.
- Gottlieb, P. D., & Joseph, G. (2006). College - to - work Migration of Technology Graduates and Holders of Doctorates within the United States. *Journal of Regional Science*, 46(4), 627-659.
- Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, 60(1), 126-142.
- Kodrzycki, Y. K. (2001). Migration of Recent College Graduates: Evidence from the National Longitudinal Survey of Youth. *New England Economic Review*, 1(1), 13-34.
- Kyung, W. (1992). *Determinants of In-migration of College Students to the State of New York*. (Center for Urban Studies Publications and Reports Paper 40) Retrieved from Portland State University website http://pdxscholar.library.pdx.edu/cus_pubs/40
- Ishitani, T. T. (2011). The Determinants of Out-migration Among In-state College Students in the United States. *Research in Higher Education*, 52(2), 107-122.
- Levy, M. B., & Wadycki, W. J. (1974). Education and the Decision to Migrate: An Econometric Analysis of Migration in Venezuela. *Econometrica: Journal of the Econometric Society*, 42(2), 377-388.

- Li, F., Liu, F., & Guo, Z. (2009). An Empirical Study of Migration in Postgraduate Employment. *Tsinghua Journal of Education*, 30(4), 67-71.
- Liu, Y., & Shen, J. (2014). Spatial Patterns and Determinants of Skilled Internal Migration in China, 2000–2005. *Papers in Regional Science*, 93(4), 749-771.
- Liu, Y., Shen, J., Xu, W., & Wang, G. (2017). From School to University to Work: Migration of Highly Educated Youths in China. *The Annals of Regional Science*, 59(3), 651-676.
- Machin, S., Salvanes, K. G., & Pelkonen, P. (2012). Education and Mobility. *Journal of the European Economic Association*, 10(2), 417-450.
- Mak, J., & Moncur, J. E. (2003). Interstate Migration of College Freshmen. *The Annals of Regional Science*, 37(4), 603-612.
- Malamud, O., & Wozniak, A. (2012). The Impact of College on Migration Evidence from the Vietnam Generation. *Journal of Human Resources*, 47(4), 913-950.
- Mcfadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*; Zarembka, P., Ed.; Academic Press: New York, NY, USA, 1974; pp. 105–142.
- McHenry, P. (2013). The Relationship between Schooling and Migration: Evidence from Compulsory Schooling Laws. *Economics of Education Review*, 35, 24-40.
- Mincer, J. (1978). Family Migration Decisions. *Journal of Political Economy*, 86(5), 749-773.

- Mixon Jr, F. G., & Hsing, Y. (1994). The Determinants of Out-of-state Enrollments in higher Education: A Tobit Analysis. *Economics of Education Review*, 13(4), 329-335.
- Mueser, P. R., & Graves, P. E. (1995). Examining the Role of Economic Opportunity and Amenities in Explaining Population Redistribution. *Journal of Urban Economics*, 37(1), 1-25.
- Newbold, K. B. (2001). Measuring Internal Migration Among the Foreign-born: Insights from Canadian Data. *Review of Regional Studies*, 31(2), 177-195.
- Parikh, A., & Van Leuvensteijn, M. (2003). Internal Migration in Regions of Germany: A Panel Data Analysis. *Applied Economics Quarterly*, 49(2), 173-192.
- Plane, D. A., & Bitter, C. (1997). The Role of Migration Research In Regional Science¹. *Papers in Regional Science*, 76(2), 133-153.
- Rogers, A., Raquillet, R., & Castro, L. J. (1978). Model Migration Schedules and Their Applications. *Environment and Planning A*, 10(5), 475-502.
- Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2), 135-146.
- Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, 51(1), 1-17.
- Schwartz, A. (1973). Interpreting the Effect of Distance on Migration. *Journal of Political Economy*, 81(5), 1153-1169.

Schwartz, A. (1976). Migration, Age, and Education. *Journal of Political Economy*, 84(4, Part 1), 701-719.

Shen, J., & Liu, Y. (2016). Skilled and Less-skilled Interregional Migration in China: A Comparative Analysis of Spatial Patterns and the Decision to Migrate in 2000–2005. *Habitat International*, 57, 1-10.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70(5, Part 2), 80-93.

Tornatsky, L. G. (2001). *Who Will Stay and who Will Leave?: Individual, Institutional and State-level Predictors of State Retention of Recent Science and Engineering Graduates, Policy and Research Recommendations*. Southern Growth Policies Board.

Spilimbergo, A., & Hanson, G. H. (1996). *Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the US-Mexico Border* (NBER Working Paper No. 5592). Retrieved from National Bureau of Economic Research website: <https://www.nber.org/papers/w5592>.

Stark, O., & Bloom, D. E. (1985). The New Economics of Labor Migration. *The American Economic Review*, 75(2), 173-178.

Steahr, T. E., & Lowe, R. A. (1975). *Patterns and Trends of College and University Student Migration in the United States*. (Storrs Agricultural Experiment Station Paper 54), Retrieved from the University of Connecticut website. <https://opencommons.uconn.edu/saes/54>.

- Tuckman, H. P. (1970). Determinants of College Student Migration. *Southern Economic Journal*, 184-189.
- Van der Gaag, N., van Wissen, L., Rees, P., Stillwell, J., & Kupiszewski, M. (2003). Study of Past and Future Interregional Migration Trends and Patterns Within European Union Countries: In Search of a Generally Applicable Explanatory Model, Eurostat Report 2002/S 67-052015/EN.
- Venhorst, V., Van Dijk, J., & Van Wissen, L. (2011). An Analysis of Trends in Spatial Mobility of Dutch Graduates. *Spatial Economic Analysis*, 6(1), 57-82.
- Wolpert, J. (1965). Behavioral Aspects of the Decision to Migrate. *Papers of the Regional Science Association*, 15(1), 159-169.
- Wozniak, A. (2010). Are College Graduates More Responsive to Distant Labor Market Opportunities? *Journal of Human Resources*, 45(4), 944-970.
- Young, A. (2013). Inequality, the Urban-rural Gap, and Migration. *The Quarterly Journal of Economics*, 128(4), 1727-1785.
- Yousefi, M., & Rives, J. (1987). Migration Behavior of College Graduates: An Empirical Analysis. *Journal of Behavioral Economics*, 16(3), 35-49.
- Yue, C. J., & Zhou, J. B. (2005). Why Do the College Graduates Choose Inter-province Employment. *Tsinghua Journal of Education*, 2, 34-41.
- Yue, D. (2010). Factors Affecting Employment Opportunities for College Graduates. *China Higher Education Research*, 11, 68-71.

Zheng, Z. & Yang, K. (2013). (In Chinese). Current Pattern and Future Trend of China's Migration. *People's Tribune*, 2013 (4): 6-9.

VITA

Zixin Liu was born and grew up in Liuzhou, Guangxi, China. She received her bachelor's degree in finance in July 2013 from the Beijing Language and Culture University. She started her graduate study in economics at the University of Missouri-Columbia in 2013 under the supervision of Dr. Peter Mueser. Zixin's research focuses on labor economics, economics of education, and applied econometrics. She received her master's degree in economics and Ph.D. degree in economics in May 2021 from the University of Missouri-Columbia.