2018

THREE ESSAYS ON COLLEGE EARNINGS PREMIUM AND CHINA'S HIGHER EDUCATION EXPANSION

Chenxu Hu
University of Kentucky, chenxu.hu@uky.edu
Digital Object Identifier: https://doi.org/10.13023/etd.2018.319

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Chenxu Hu, Student

Dr. Chris Bollinger, Major Professor

Dr. Josh Ederington, Director of Graduate Studies

THREE ESSAYS ON COLLEGE EARNINGS PREMIUM AND CHINA'S HIGHER EDUCATION EXPANSION

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Business and
Economics at the University of
Kentucky

By
Chenxu Hu
Lexington, Kentucky

Director: Dr. Chris Bollinger, Professor of Economics
Lexington, Kentucky 2018

ABSTRACT OF DISSERTATION

THREE ESSAYS ON COLLEGE EARNINGS PREMIUM AND CHINA'S HIGHER EDUCATION EXPANSION

My dissertation consists of three essays that study the college premium in China and how it has been affected by China's higher education expansion.

In the first essay, I utilize the high education expansion as exogenous source to estimate the college premium. The rapidly changing access to college provides a rare opportunity to estimate a local treatment effect (LATE) of college education on earnings by utilizing the drastic increase in college admission rate in 1999. I also utilize the yearly admission rate as an instrumental variable for the endogenous college education. Using China Household Income Project 2013, the two IV estimates of college premium are 75.7 and 57.5 log points respectively.

The second essay examines the trends of the college earnings premium by age groups from 1995 to 2013 in China. Specifically, based on China Household Income Projects, the college premium for the younger group (age 25-34) stagnated, while the college premium for the older group (age 45-54) increased substantially. I attribute the stagnation for the younger group to the fast-growing relative supply of younger college workers due to China's higher education expansion. Holding the age cohort and survey year constant, a one unit increase in log relative size of college workers leads to 10.3 log points decrease in college premium.

The third essay further explores the channel through which the cohort size affects the college premium. Using Blinder-Oaxaca decomposition, I find that, for all survey years and age groups, the differential of the higher-skilled occupations share between college and non-college educated workers only explains a small part of college premium, 10%-30%. The part due to the higher-skilled occupational premium is negligible. Over 70% of the college premium is contributed by the college premium among the workers with lower-skilled occupations.

KEYWORDS: College Premium; Higher Education Expansion; China; Cohort Size; Occupation

Author's signature: Chenxu Hu

Date: July 26, 2018

THREE ESSAYS ON COLLEGE EARNINGS PREMIUM AND CHINA'S HIGHER EDUCATION EXPANSION

By
Chenxu Hu

Director of Dissertation: Dr. Chris Bollinger

Director of Graduate Studies: Dr. Josh Ederington

Date: July 26, 2018

To my wife Dan, my father Xingbo, and my mother Yuqiu.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support of several individuals. First, I wish to thank my chair, Dr. Chris Bollinger, for patiently giving of his time, insight and guidance in all stages of the dissertation. Next, I thank Drs. James Ziliak, Olga Malkova, and Xin Ma, who form the remainder of my dissertation committee, for giving me valuable input. I am thankful to Dr. Yuqing Zheng who agreed to serve as an outside examiner to my dissertation defense. Thanks to Jeannie Graves and Debbie Wheeler at the Department of Economics for all their help and support. I sincerely thank my nice roommates Hao Guo and Zheng Yang for being great help since my second year at Kentucky. I thank Nicholas Moellman for his proofreading.

I am grateful to the Gatton Doctoral Fellowship and Chinese Scholar Council Scholarship for providing me with financial support throughout my studies at Kentucky, without which I might not have continued. Finally, I thank my wife Dan, my parents Xingbo and Yuqiu for their love and support.

TABLE OF CONTENTS

Acknowledgments . . . iii

Table of Contents . . . iv

List of Tables . . . v

List of Figures . . . vi

Chapter 1 Introduction . . . 1

Chapter 2 Earnings Effect of College Education: IV Estimates Based on China's Higher Education Expansion . . . 5
- 2.1 Introduction . . . 5
- 2.2 Literature Review . . . 7
- 2.3 Background . . . 9
- 2.4 Data Description . . . 11
- 2.5 IV Estimation Based on 1999 Expansion . . . 15
- 2.6 IV Estimation Based on Admission Rates . . . 19
- 2.7 Conclusion . . . 23

Chapter 3 Effects of Cohort Size on College Premium: Evidence from China's Higher Education Expansion . . . 41
- 3.1 Introduction . . . 41
- 3.2 Theoretical Framework . . . 44
- 3.3 Empirical Approach . . . 46
- 3.4 Data . . . 51
- 3.5 Results . . . 55
- 3.6 Robustness Checks . . . 57
- 3.7 Conclusion . . . 59

Chapter 4 Education-Occupation Match and the Trends in the College Earnings Premium by Age Groups in Urban China, 1995-2013 . . . 77
- 4.1 Introduction . . . 77
- 4.2 Data Description . . . 78
- 4.3 Decomposition Analysis . . . 80
- 4.4 Cohort Size Effect on College Premium by Occupation Categories . . . 83
- 4.5 Conclusion . . . 86

Bibliography . . . 102

Vita . . . 107

LIST OF TABLES

2.1 Summary Statistics 33
2.2 Summary Statistics for the Sample Including all College and High School Graduates 34
2.3 Validity Test for Joint Significance of Baseline Covariates 35
2.4 First Stage and Reduce Form Estimations 36
2.5 OLS and RD-IV Estimates of College and 4-year College Premiums . . . 37
2.6 Estimated Effects of IVs on Years of Education, College Attendance and Log Annual Earnings 38
2.7 OLS and IV Estimates of Rate of Returns to Higher Education 39
2.8 OLS and IV College Premium Estimates 40

3.1 Summary Statistics: Male Workers Only 68
3.2 Birth Year Fixed Effects on College Premium and Relative Size 69
3.3 Basic Estimates for Effects of Age Specific Relative Size of College Workers on College Premiums 70
3.4 Heterogeneous Relative Size Effects across Age Groups 71
3.5 The Results to Sample including only High-School and 4-Year College Workers 72
3.6 Testing Assumption: Identical Elasticity of Substitution for College and Non-College Workers 73
3.7 Results Using Individual Data Controlling for Province, Occupation and Industry 74
3.8 Robustness of The Results to Female Sample and Pooled Sample Including both Male and Female 75
3.9 Robustness of The Results to Several Alternative Specifications 76

4.1 List of Approaches to Determining b^* 95
4.2 Summary Statistics: Male Workers Only in CHIP Urban Data 96
4.3 Summary Statistics: Higher-Skilled Occ. Distribution by Education, Year, and Age Group 97
4.4 Summary Statistics: High-Skill Occ. Distribution by Education, Year, and Age Group 98
4.5 Decomposition Results by Oaxaca-Blinder Decomposition 99
4.6 Basic Estimates for Effects of Age Specific Relative Size of College Workers on College Premiums: Lower-Skilled Occupations 100
4.7 Basic Estimates for Effects of Age Specific Relative Size of College Workers on College Premiums: Higher-Skilled Occupations 101

LIST OF FIGURES

2.1 Changes in China's wage determination 1988-2009 25
2.2 China's Higher Education Expansion:Scales 26
2.3 China's Higher Education Expansion:Rates 26
2.4 Distribution of the Age When Taking the NCEE 27
2.5 Cohort-Specific NCEE-Age Population, Number and Ratio of NCEE-Takers 28
2.6 Cohort-Specific Shares of College Graduates of Sample and Population . 29
2.7 Test of the discontinuity of college attendance and log annual earnings at cutoff cohort 1999 30
2.8 Test of the discontinuity of 4-year college attendance and log annual earnings at cutoff cohort 1999 31
2.9 NCEEHukou-specific Local Averages of Education and Earnings 32

3.1 Trends of College Premium and Relative Supply of College Workers by Age Groups: China 61
3.2 Trends of College Premium and Relative Supply of College Workers by Age Groups: The U.S. 62
3.3 Trends of College Premium and Relative Supply of College Workers by Age Groups: Japan 63
3.4 Age-Year Cell Specific Log Relative Sizes and Estimated College Premiums 64
3.5 Male Workers' Age Profiles of the College Premium Across Years 65
3.6 Male Workers' Age Profiles of Relative Size Across Years 65
3.7 Demographical Change and Higher Education Expansion in China 66
3.8 Birth Year Profiles of the Share of College Workers 66
3.9 Predicted Birth Group Fixed Effects on the College Premium 67
3.10 Predicted Birth Group Fixed Effects on the Share of College Workers . . 67

4.1 Occupation Categories and Education levels 89
4.2 Trends in the College Premium across Age Groups 90
4.3 Trends in the Explained Earnings Gap across Age Groups 90
4.4 Trends in the Unexplained Earnings Gap Due to Occupational Premiums across Age Groups 91
4.5 Trends in the Unexplained Earnings Gap Due to the Constants across Age Groups 91
4.6 Trends in the Shares (Differential) of Higher-Skilled Occupations by Education and Age Groups 92
4.7 Trends in the Percentages of the Explained Earnings Gap By Age Groups 93
4.8 Trends in the Percentages of the Unexplained Earnings Gap Due to Occupational Premiums By Age Groups 93
4.9 Trends in the Percentages of the Unexplained Earnings Gap Due to Constants By Age Groups 94

Chapter 1 Introduction

It has been confirmed that individuals with higher education level, on average, earn higher wages when holding other observed personal characteristics constant. But the challenging question is to what extent are higher earnings are caused by higher education levels. The OLS estimation of the returns to education are suspected to be overestimated due to unobserved personal characteristics (eg. ability) both positively affecting education attainment and earnings. In contrast, potential measurement error in education years or levels tend to attenuate the OLS estimates. Therefore, simple OLS estimates might not be satisfactory answers to the question above. The question has become especially relevant for countries with dramatic rise in the scale of higher education enrollment in the last two decades.[1] Whether the large scale of investments in higher education pay off in the future depends on the raised earnings caused by college education.

In chapter 2, I use China's higher education expansion to identify the causal earnings effect of college education relative to high school level.[2] China's higher education has expanded substantially since 1977 when the national college entrance examination(NCEE) restored. In 1999, the Chinese government launched an ambitious expansion in higher education. The nationwide college admission rate increased from about 34% in 1998 to 56% and the admission rates remained stable at high levels around 60% in the following years till 2008. Individuals taking the NCEE in 1999 had significantly higher probability to be admitted than those in 1998. First, I make use of the tremendous exogenous expansion in 1999 to identify the earnings effect of college education. The government-controlled college enrollment scale and demographically affected college applicants scale result in a plausibly exogenous, or predetermined at least, variation of the cohort-specific college admission rates. And the college admission rates were sufficiently various for making identification. So, I also implement a different IV estimation strategy using the cohort-specific college admission rate as instrumental variable. Using the individual data drawn from China Household Income Project (CHIP) which is a widely used repeated cross-sectional survey data to study China's labor market and households, the primary estimate of the earnings effect of college education is 75.7 log points, and the effect of 4-year college education is 92.1 log points. These estimated returns have very strong policy implication that the 1999 expansion has indeed substantially benefited those people who were admitted into college in 1999 due to the expansion. But this is also a limitation for making inferences more generally. Our estimates by the second method reduce the limitation substantially, because the cohort-specific college admission rates

[1] See Machin and McNally (2007) for surveys of international evidence of tertiary education expansion, and Maurin and McNally (2008), Lemieux and Card (2001) for brief statements about the expansion in Britain, France and Canada.

[2] In this paper, we define college as both 3-year and 4-year college education. The specific earning effect of 4-year college will also be discussed. We refer these effects as college premium in this paper.

affect individuals from all cohorts instead of only those of cohort 1999. Returns to one additional year of higher education is estimated at 16.1 log points and returns to college education is estimated at 57.5 log points. Our estimates demonstrate that the higher expansion policy has indeed substantially improved the earnings of those who obtained college education due to the expansion. Our study also contributes to the emerging literature applying regression discontinuity design to estimate educational returns (Oreopoulos, 2006; Fan et al., 2010) as well as the literature exploiting natural experiments as instrumental variables to identify the earnings effect of education (Lemieux and Card, 2001; DuFLo, 2001; Maurin and McNally, 2008; Giles et al., 2015).

In chapter 3, I further examine how the expansion affects the college earnings premium. As a leading proximate cause of rising overall earnings inequality since the 1980s in the U.S., the increase in the college/high school wage premium has been well documented. Authors such as Katz and Murphy (1992), Acemoglu (2002), and Autor et al. (2008) have explained the rise as the consequence of an accelerated rise in the relative demand for college graduates and an abrupt slowdown in the growth of the relative supply of college graduates.[3] These studies focus on the aggregate trend of the college wage premium that may conceal independent trends by age groups. Card and Lemieux (2001) argue that heterogeneous trends of college premium by age groups may arise if workers in different age groups within the same education group are imperfectly substitutable and the trends of the relative supply of college workers are heterogeneous by age groups. Using data from the United States, the United Kingdom and Canada, they demonstrate the imperfect substitution between age groups and attribute the observed relative rise in the college premium for younger workers since the early or mid 1980s to the stagnated growth of the relative supply of college educated workers among the young during the same periods.[4] However, little evidence from other countries has been added until recently. Kawaguchi and Mori (2016) reveal the heterogeneous trends of the college premium by age groups between 1986 and 2008 in Japan. This chapter adds evidence to this literature by documenting the divergent trends of college premium by age groups between 1995 and 2013 in China, and examines how the college premium is affected by the age group specific relative size of college educated workers.[5]

Using China Household Income Project (CHIP) 1995, 1999, 2002, 2007, and 2013, five repeated cross-sectional surveys, I find that the trends of the college premium between 1995 and 2013 by age groups are substantially different. In figure 3.1(a), the college premium as measured by log earnings ratio was very similar for younger

[3]It is argued that the increase may have been driven by both skill-biased technological change (SBTC) featured by the computer revolution and the outsourcing of manufacturing. Katz et al. (1999) and Autor et al. (2008) support the idea of SBTC, and Feenstra and Hanson (2001) support the idea of outsourcing. The growth of college graduation rates stagnated for cohorts born in the early 1950s and entered labor market in late 1970s. See Card and Lemieux (2001) for details.

[4]The relative rise in college premium for younger workers commenced 5 years later in the U.K. and Canada than in the U.S.

[5]Considering that there exists certain amount of workers below high school education in China,

(age 25-34) and older (age 45-54) groups, about 25 percentage points in 1995. As of 2013, the college premium for the younger group was about 30 percentage points, similar to the level in 1995, while the college premium for the older group was about 50 percentage points, nearly double that of 1995. In figure 3.1(b), we present the age group specific trends of the relative supply of college workers measured as log employment ratio. The relative supply for the younger group increased substantially while that for the older group was quite stable during the same period. Comparing these two figures, the stagnation of the college premium for the younger group between 1995 and 2013 was potentially due to the fast growing relative supply of college workers. Figures 3.2 and 3.3 show that in the U.S. and Japan, unlike in China, the college premium for the older group decreased with respect to the younger group while the relative supply for the older group increased with respect to the younger group.[6] If technological progress positively affects the college premium for the younger group particularly as the literature argues, the negative age group specific supply effects will be overestimated for the U.S. and Japan, and underestimated for China. I follow the empirical strategy by Card and Lemieux (2001) to construct the college premium and relative supply by age and survey year, and to further regress the cell-specific college premium against the relative supply. The supply effect on the college premium is estimated to be about -0.1 by our main specification. That implies, when holding the age cohort and survey year constant, a one unit increase in the log relative size of college workers is associated with about 10 percentage points decrease in the college premium. The more comparable result, by focusing on the college/high school earnings premium, is about -0.18 which is slightly lower than -0.2 in the U.S. and -0.23 in the U.K. while almost same as the results for Japan and Canada. That the negative supply effect in China is so close to the other four countries is remarkable in view of the very different economic development levels, trends of the college premiums and the relative supply, and higher education expansion phases between China and the other four countries.

The previous studies attribute the divergent trends in college premiums across age groups to the negative cohort size effect. As the relative supply of college educated workers increases, in a competitive labor market, college workers' average relative earnings which equal the relative marginal product decrease. However, no studies have examined the channel through which the cohort relative supply affect the college premium. China's higher education expansion has led to an increasing number of college-educated workers, especially among the younger cohorts. More and more young workers can be placed into occupations which have typically been held by people with less education and lower earnings. It's possible that the cohort relative supply of college educated workers negatively affects the college premium by rematching workers with different education levels to different occupations. Chapter 4 examines how much the education-occupation match accounts for college earnings

I focus on the college premium with respect to non-college workers. Results for the college/high school premium will also be discussed and compared with existing studies.

[6] These two figures are taken from the paper by Kawaguchi and Mori (2016) who compare the

premium and to what extent the trends in education-occupation match contribute to the divergent trends in college earnings premium across age groups in urban China between 1995 and 2013. Using Blinder-Oaxaca decomposition (Oaxaca, 1973; Blinder, 1973), it's found that, for all survey years and age groups, the differences in the share of higher-skilled occupations between college and non-college educated workers only explains a small part of college premium, 10%-30%. The part due to the occupational premium is negligible. Over 70% of the college premium is contributed by the college premium for the workers with lower-skilled occupations. These unexpected results reveal that the relative supply of college educated workers among certain age groups has a general effect rather than age group specific effect on the occupation reallocation.

The rest of this paper is followed by the three essays, chapters 2, 3, and 4.

---

trends between the U.S. and Japan.

Chapter 2 Earnings Effect of College Education: IV Estimates Based on China's Higher Education Expansion

2.1 Introduction

Hundreds of studies have revealed the importance of education in modern labor markets. The question of how education affects earnings received overwhelming interest. It has been confirmed that individuals with higher education level, on average, earn higher wages when holding other observed personal characteristics constant. But the challenging question is to what extent are higher earnings are caused by higher education levels. The OLS estimation of the returns to education are suspected to be overestimated due to unobserved personal characteristics (eg. ability) both positively affecting education attainment and earnings. In contrast, potential measurement error in education years or levels tend to attenuate the OLS estimates. Therefore, simple OLS estimates might not be satisfactory answers to the question above. The question has become especially relevant for countries with dramatic rise in the scale of higher education enrollment in the last two decades.[1] Whether the large scale of investments in higher education pay off in the future depends on the raised earnings caused by college education.

This paper uses China's higher education expansion to identify the earnings effect of college education relative to high school level.[2] China's higher education has expanded substantially since 1977 when the national college entrance examination(NCEE) restored. In 1999, the Chinese government launched an ambitious expansion in higher education. The nationwide college admission rate increased from about 34% in 1998 to 56% and the admission rates remained stable at high levels around 60% in the following years till 2008. Individuals taking the NCEE in 1999 had significantly higher probability to be admitted than those in 1998. First, we make use of the tremendous exogenous expansion in 1999 to identify the earnings effect of college education. This strategy is reliable if we believe the earnings effect of the unobservables was smooth along with the individual's NCEE-taking year. Then, if we control for the smooth earnings effect of the unobservables using a low-order polynomial, the earnings differential between cohorts 1998 and 1999 will only be due to the difference in share of college educated workers between the two cohorts.[3] This is a special case of the regression discontinuity design with discrete assignment variable, the NCEE year.[4] The estimation can be implemented by the standard IV estimation

[1]See Machin and McNally (2007) for surveys of international evidence of tertiary education expansion, and Maurin and McNally (2008), Lemieux and Card (2001) for brief statements about the expansion in Britain, France and Canada.

[2]In this paper, we define college as both 3-year and 4-year college education. The specific earning effect of 4-year college will also be discussed. We refer these effects as college premium in this paper.

[3]In this paper, a cohort refers to a group of individuals who took the NCEE in a same year.

[4]See Lee and Lemieuxa (2010),Lee and Card (2008) for details.

procedures. The identified college earnings effect is the local average treatment effect (LATE) of college education on earnings for the subpopulation induced into college by the drastic expansion in 1999.

The government-controlled college enrollment scale and demographically affected college applicants scale result in a plausibly exogenous, or predetermined at least, variation of the cohort-specific college admission rates. And the college admission rates were sufficiently various for making identification. So, we implement a different IV estimation strategy using the cohort-specific college admission rate as instrumental variable. However, the same identification issue may arise if there exists cohort fixed effect which is not sufficiently smooth. To deal with this issue, we construct interaction term between college admission rate and residence status when took the NCEE (NCEE-Hukou) as additional instrumental variable with which we can control for the cohort fixed effect in the specification.[5]

In this paper, the individual data is drawn from China Household Income Project (CHIP) which is a widely used repeated cross-sectional survey data to study China's labor market and households. The cohort specific variable, such as yearly NCEE admission rate, college-age population and so on, are drawn from China's Statistics Year Books. We only use the latest released 2013 wave of CHIP because it surveys the year when one took the NCEE. Therefore, it's possible to match the cohort-specific college admission rates to each individual accurately. And CHIP 2013 includes enough observations who took the NCEE after the drastic expansion in 1999. By the first estimation strategy, the primary estimate of the earnings effect of college education is 75.7 log points, and the effect of 4-year college education is 92.1 log points. These estimated returns have very strong policy implication that the 1999 expansion has indeed substantially benefited those people who were admitted into college in 1999 due to the expansion. But this is also a limitation for making inferences more generally. Our estimates by the second method reduce the limitation substantially, because the two instruments used, cohort-specific college admission rate and the interaction between the admission rate and NCEE-Hukou status, affect individuals from all cohorts instead of only those of cohort 1999. Returns to one additional year of higher education is estimated at 16.1 log points and returns to college education is estimated at 57.5 log points.

As typically found in the literature (Card, 1999b; Heckman et al., 2006), our IV estimates are larger than OLS estimates. This demonstrate that the ability bias may not be important and the measurement error in reported education levels leads to downward biased OLS estimates. Most studies using Chinese data also find large IV estimates than OLS estimates except Li et al. (2012) and Giles et al. (2015). Li et al. (2012) use twins data and follow the method by Ashenfelter and Krueger (1994),

[5]Hukou refers to China's residence registration system. People born in rural area are usually registered as rural Hukou and those born in urban area are registered as urban Hukou. The Hukou status can be changed from rural to urban. Thus, in our data, people in the urban sample may used to be registered as rural Hukou. NCEE-Hukou is defined as people's Hukou status when they took the NCEE. By this variable, we differentiate the places where people get their education before taking the NCEE. The urban status for NCEE-Hukou is different from the urban status for being

obtained the rate of return to education, 8 percent by OLS and 3.8 percent by FE-IV estimation. Giles et al. (2015) utilize the variation in disruptions to education due to the Cultural Revolution as instrumental variable for schooling and obtained 9.8 percent by OLS estimation and 8 percent by IV estimation. These two studies reveal that the ability bias is important in China. Li et al. (2012) provide an explanation that the exam-oriented high school education acts as a selection tool providing no knowledge or training that will enhance earnings. Thus, their IV estimates of the overall return to years of education are lower than OLS estimates. The studies focusing on the return to college education with Chinese data all find larger IV estimates (Fan et al., 2010; Wang, 2012; Wang et al., 2014).

Our estimates demonstrate that the higher expansion policy has indeed substantially improved the earnings of those who obtained college education due to the expansion. Our study also contributes to the emerging literature applying regression discontinuity design to estimate educational returns (Oreopoulos, 2006; Fan et al., 2010) as well as the literature exploiting natural experiments as instrumental variables to identify the earnings effect of education (Lemieux and Card, 2001; DuFLo, 2001; Maurin and McNally, 2008; Giles et al., 2015).

This paper proceeds in the following order. Section 2 reviews the literature. Section 3 introduces China's national college entrance selection mechanism and China's higher education expansion since 1977. Section 4 describes the data. Sections 5 and 6 discuss the two sets of IV estimates respectively. The last section concludes.

2.2 Literature Review

Identification strategies for the causal effect of education on earnings

Identifying the economic returns to education is a huge literature. The base model used was popularized and estimated by Mincer (1974) regressing earnings or wages on schooling, quadratic experience, and other personal characteristics as control variables. But since Becker (1964), a major concern has been that basic OLS estimates might overstate the true effect of education on earnings due to unobserved ability which positively affecting both schooling attainment and earnings. Additionally, the issue of measurement error in self-reported schooling attainment has been argued to bias the OLS estimates downward.

Early attempts controlled for ability test scores directly to test the existence of ability bias and revealed that the effect of measured ability on earnings varies across ages and education levels. (e.g. Taubman and Wales (1974)). Ashenfelter and Krueger (1994) use twins data to control unobserved ability or family background factors, and used sibling-reported schooling as IV to correct measurement error bias. They pointed out that ability bias is actually ignorable while the measurement error biases are potentially important.

Instrumental variable estimation is the most common identification strategy. A list of commonly used instrumental variables for schooling were documented by Card

(1999b) and Heckman et al. (2006): family background variables (Card, 1993; Miller et al., 1995; Taber, 2001), geography location (Card, 1993; Kling, 2001), tuition (Kane and Rouse, 1995), local labor market variables (Carneiro and Heckman, 2002), quarter of birth (Angrist and Krueger, 1991). Most of the IV estimates are larger than OLS estimates and with larger variances. Some other studies exploited exogenous variation from policy experiments. DuFLo (2001) used the variation in the regional primary school constructions during 1970s in Indonesia. Lemieux and Card (2001) used the regional difference in a Canadian policy facilitating returning veterans pursuing college education after WWII. Maurin and McNally (2008) utilized the lowered college entrance threshold in 1968 due to the students protest.

There is a large literature applying Regression Discontinuity design to evaluate treatment effect[6]. However, few applications exist to the estimation of returns to education. The first RD application by Oreopoulos (2006) utilized the changes to the minimum school-leaving age in the United Kingdom in 1947 and 1957. Another study by Fan et al. (2010) utilize the unique feature of Chinese college admission mechanism, a strict test score-based admission rule.

Returns to education in China

Returns to education in Chinese urban labor market in the 1980s and 1990s were extremely low. Byron and Manaloto (1990), Johnson and Chow (1997) and Liu (1998) all report quite low rates of return to one additional year of schooling, between 3 and 4 percent. These estimates are much lower than the world average level, 10.1 per cent (Psacharopoulos, 1994). The returns have increased since the mid-1990s. The best set of consistent estimates over time uses repeated cross-sectional urban survey data collected by the National Bureau of Statistics from 6 provinces in different regions from 1988 to 2001 and finds that over this period the increase in annual wages in urban China associated with an additional year of schooling grew from 4 percent to over 10 percent (Zhang et al., 2005). Ge and Yang (2011) also find that the rate of return to one additional year of schooling in urban China, estimated using OLS, increased from 3.6 per cent in 1988 to 11.4 per cent in 2007. Meng (2012) extends the data set to 2009 and update more recent trends in China's wage structure. Figure 2.1 presents the changing effects of observable characteristics on the urban wage structure between 1988 and 2009. The reason for the low rates of return to education in the pre-reform period and initial stages of the post-reform period is that China's long-term allocation of labor resulted in a relatively equal distribution of income (Gao and Smyth, 2015).[7]

Simple OLS estimates of the returns in urban China are subject to two potential bias: measurement error in self-reported schooling attainment and unobservable ability. Li et al. (2012) collect twins data from five Chinese cities in 2002 and follow

---

in urban sample.

[6] See Lee and Lemieuxa (2010) for surveys of the recent applications of RD design in a lot of topics.

[7] Urban Household Survey are not public data. So here I cite the figure 2 in Meng (2012) to offer an overview of the change in China's wage determination.

the empirical methods by Ashenfelter and Krueger (1994). The estimated rates of return to one additional year of schooling in term of log monthly wages are only 2.7 percent, and 3.8 percent after using sibling reports as an instrument to correct for measurement error in self-reported schooling attainment. In contrast to the basic OLS estimate, 8 percent, they reveal a much greater degree of ability bias in China than that found by Ashenfelter and Krueger (1994) with U.S data. This substantial ability bias is attributed to China's test-oriented education system for high school level and below. Giles et al. (2015) use individual level variation in the extent of city-wide disruptions to education due to the Cultural Revolution as an instrumental variable for schooling. The IV estimate of the rate of return in term of hourly wages in urban china in 2001 is 8 percent compared to 9.6 percent for OLS estimate. Wang et al. (2014) follow in the semi-parametric estimation framework surveyed by Heckman et al. (2006), find that the 4-year college premium increased from an insignificant 24.4% in 1988, to an insginificant 42% in 1995, and then to a very significant 165.1% in 2002.

Impacts of China's higher education expansion on labor market outcomes

Meng (2012) finds that basic OLS estimates of returns to college-and-above education (with respect to primary school level) rise from around 16 percent in late 1980s to over 50 percent by 2003, but since then the returns have slipped back slightly. It is argued to be related to the large influx of graduates due to the higher education expansion in 1999. Li et al. (2014) use DID strategy with nationally representative population surveys from 2000 and 2005, treat the expansion in 1999 as a policy shock , then find that China's higher education since 1999 has sharply increased the unemployment rate among young college graduates, by about 9 percent.

2.3 Background

College Entrance Selection

Due to the Cultural Revolution from 1966 to 1976, China's national college entrance examination (NCEE) was suspended for a decade until 1977. Since then, the NCEE held every year nationwide has been the only formal channel for high school graduates to enter college. The NCEE is held at the end of the spring semester every year. High school graduates across the country take the examination simultaneously over a three-day period. Before 2003, it was held in July, but has been moved to June since 2003.

The NCEE is not uniform across the country, but uniform within each province and each direct-controlled municipality.[8] The NCEE has three mandatory subjects in every province: Chinese, Mathematics, and a foreign language (usually English, but this may also be substituted by Japanese, Russian or French). The other six standard subjects are three sciences, Physics, Chemistry and Biology; and three humanities,

[8] Four direct-controlled municipalities: Beijing, Shanghai, Tianjin and Chongqing.

History, Geography and Political Education. Applicants to science/engineering or art/humanities programs typically take one to three from the respective category. Since the 2000s, the science integrated test, the humanities integrated test or whole integrated test has been introduced in some provinces.

Every student receives a score in the NCEE and the score is essentially the only criterion for admission to higher education, in contrast to the US and other OECE countries where application requirements are based on SAT scores, recommendation letters from teachers, participation in extracurricular activities, and high school grades. In different provinces and across different years, students were required to apply for their intended university or college prior to the exam, after the exam, or more recently, after they learned of their scores, by filling a list of ordered preferences. The application list is classified into several tiers (early admissions, key universities, regular universities, vocational colleges)[9], each of which can include around 4-6 intended choices in institutions, though typically an institution would only admit students who apply to it as their first choice in each tier. Each province has a set of score thresholds for these different tiers and the threshold for vocational college is also the bottom line for college admission. Students failing in the first attempt can make another one in the following year and there is no age limitation.

A university usually sets different admission quota across provinces, with a higher number for its home province. The higher education resources are distributed unevenly across China, so it is argued that people are being discriminated during the admission process based on their geographic region.[10] The unequal admission for different provinces intensifies competition among students from provinces with fewer higher education resources.

The Expansion

Figure 2.2 and 2.3 depict China's higher education expansion from 1977 to 2012 by enrollment levels and rates respectively. In figure 2.2, the number of NCEE takers in 1977 and 1978 were about 6 million which was twice more than the average number between 1983 to 1998. The larger number of NCEE takers in 1977 and 1978 were to address the fact that the NCEE was suspended for ten years from 1966-1976 during the Cultural Revolution. Those high school graduates who wished to pursue higher education but had no channel during the Cultural Revolution took the NCEE right after it was restored.

The number of NCEE takers dropped rapidly to 1.67 million in 1983 due to the quite low enrollment rate, only 4.74% in 1977. Meanwhile the overall enrollment was growing gradually from 0.27 million in 1977 to 0.62 million in 1985. The decreasing NCEE takers paired with the growing college enrollment rapidly increased the enrollment rate which is shown in figure 2.3. The enrollment rate increased from only

[9]Early admissions are mainly military institutions. Key university and regular university are 4-year while vocational college is 3-year.

[10]For example, compared to Beijing, Henan province has fewer universities per capita. Therefore, Henan province usually receives fewer admission slots compared with Beijing, hence Henan province

4.74% in 1977 to 35.17% in 1985.

Since 1986, not only the overall college enrollment but also the 3-year(vocational college awarding vocational college degree) and 4-year(university awarding bachelor degree) college enrollment have been reported in China statistical yearbook. From 1986 to 1998, the number of NCEE takers fluctuated around 2.8 million. Meanwhile the enrollment of 3-year college increased steadily during the first half of this period and then started to decrease in 1994. Both overall enrollment and of 4-year college enrollment expanded steadily. Figure 2.3 shows that the overall enrollment rate fluctuated between 20% and 40% while the other two enrollment rates both fluctuated between 10% and 20% from 1986 to 1998.

In 1999, the Chinese government launched a tremendous expansion in higher education. Official explanations for the expansion from Mr. Lanqing Li, who was vice premier in charge of education at the time when the expansion was launched are: 1) the need for human resources to sustain the rapid development of Chinese economy; 2) government's obligation to meet the high public demand for higher education; 3) to postpone employment of high school graduates and to increase education consumption, which is an important means to stimulate domestic consumption and promote growth in related industries; 4) to discourage test-oriented teaching and learning. (Wang and Liu, 2011)

The overall enrollment jumped from 1.08 million in 1998 to 1.55 million in 1999, an increase of almost 50 percent. Thereafter, it increased to 6.89 million in 2012, reaching nearly 7 times as many as that in 1998. However, the number of NCEE takers dropped slightly in 1999 even though the expansion plan had been announced in 1998. The overall enrollment rate in 1999 jumped to 54% from 34% in 1998. Stimulated by this fact, the number of NCEE takers increased rapidly and then reached a peak of 10.5 million, in 2008. By 2012, the overall enrollment rate had increased to about 75% , the rate of 4-year college, 41% and rate of 3-year college, 34%.

2.4 Data Description

The data used in this paper are mainly obtained from China Household Income Project (CHIP) 2013 which is the latest wave. CHIP is repeating cross-sectional nationwide survey and seven waves (1988, 1995, 1999, 2002, 2007, 2008 and 2013) have been released. Gustafsson et al. (2014) comprehensively discuss the data for studying earnings, the distribution of household income and poverty in China. They point out that CHIP has taken advantage of working with the NBS in many stages of the data generating process. Households selected for the rural and urban surveys of CHIP are subsamples from the NBSâĂŹs larger surveys and cover representative provinces to make sure of the nationwide representativeness. Several facts make CHIP be the most popular survey data for studying topics about earnings in China. First, CHIP focuses on household income, provides detailed household income and family background information. Some other data sets may be less focused on income

has a lower college enrollment rate than Beijing.

information, like China Health and Nutrition Survey (CHNS) pays more attention to health information and China General Social Survey (CGSS) is basically a sociological survey. Second, there have been 5 waves of CHIP released covering the period from 1988 to 2013. This makes it possible to study the trends of earnings with CHIP. Some other data sets like China Health and Retirement Longitudinal Study (CHARLS), China Family Panel Study (CFPS), and China Household Finance Survey (CHFS) were all launched recently around 2010.

Why CHIP 2013?

Several facts make CHIP 2013 the best available data set for our research questions. First, CHIP 2013 surveys a series of questions on the (National College Entrance Examination) NCEE experiences among which 'have you ever taken the NCEE?', 'in which year did you take the NCEE as the last attempt?' and 'What was your Hukou status when taking the NCEE' are very important for our analysis. We restrict our sample to those individuals ever took the NCEE in terms of their answer to the first question. Through the reported year of the last attempt in the NCEE, we match that year's total number of NCEE-takers and college admission rate, to each individual. The NCEE-Hukou is utilized to construct one of our instrumental variables to estimate the returns to higher education.

Second, to apply the IV estimation utilizing the jump of college admission rate in 1999, we must have enough observations taking the NCEE in or later than 1999. Even if NCEE-related questions are also surveyed in CHIP 2007, the observations taking the NCEE in or later than 1999 are not enough because only four cohorts of college graduates taking the NCEE between 1999-2002 had been in labor market for at least one full year by 2007. In 2013 survey, the youngest cohort reporting labor market outcomes should be those taking the NCEE in 2008. Thus, we have 10 cohorts form 1999 to 2008 in total and they are almost half of the sample in our analysis, 1749 out of 3724 as presented in columns 3 and 6 of table 2.1.

Lastly, the oldest 1977 cohort is approaching retirement age in 2013. It implies that the samples restricted to include NCEE-cohorts 1977-2008 in this paper almost cover the whole working life cycle.[11] Thus, our estimation of the college premium is based on individuals covering almost entire working life cycle, which make the estimates more representative of the working population.

Sample Selection and Summary Statistics

Most previous studies (Zhang et al., 2005; Ge and Yang, 2011; Wang, 2012; Wang et al., 2014) only use the urban survey to estimate the college premium, given that few college graduates stay in the rural area. But since the ambitious higher education expansion in 1999, the share of college graduates in rural area has increased to a

[11] In China, the retirement age is 60 for men and 55 for women. Most people take the NCEE at 18 or 19 years old, thus the oldest 1977 cohort should be about 55 years old near retirement age.

remarkable level, 59 percent as column 5 of table 2.1 shows.[12] On the contrary, the share of high school graduates in urban area is quite low at about only 14 percent by column 4 of table 2.1. Thus, we pool the urban and rural surveys together in this paper. Another more important reason is to make sure the sample's cohort-specificshare of college graduates is as representative of the nationwide population as possible.[13,14] For the same reason, we also use both male and female samples and don't differentiate male college premium from female.

We restrict the sample to NCEE-takers between 1977 and 2008, with high school, 3-year college or 4-year college degree, between 20 and 60 years old.[15] However, using only NCEE-takers may bring in the selection issue if those NCEE-takers are significantly different from the other high school graduates who did not take the NCEE. First, our estimates may not be the population estimates because we only use a selected part of the whole high school graduates. Second, if those NCEE-takers are smarter and more likely admitted to college than the non-takers, our estimates may be underestimated. So, it is important to examine the observed differences between NCEE-takers and non-takers to see if these two groups are significantly different, even if there must be other unobserved differences. The comparison will be discussed in details in next subsection. College is defined as a binary variable, equal to one for 3- and 4-year college graduates, zero for high school graduates. By this definition, the college premium in this paper does not differentiate between 3- and 4-year college degree. We also estimate the 4-year college premium by using a subsample only including high school and 4-year college graduates. Our sample is further restricted to individuals being employed by the end of 2013 and reporting positive annual income in 2013. We use log annual earnings as the dependent variable. CHIP 2013 also surveys working months in 2013, workings days per month and working hours per day. This makes it possible to use log wage rate as alternative dependent variable.[16]

Summary statistics for our sample are presented in table 2.1. There are 3724 individuals in this sample among which 2608 are urban residents, 59 percent are males, 78 percent have college degree and 42 percent have 4-year college degree. Sample's average years of education is 14.21 and average tenure is 10.47 years.[17] Three sub-groups of cohorts 1977-1987, 1988-1998 and 1999-2008 are summarized by columns 1 to 3 respectively. Columns 4 and 5 summarize urban and rural samples. Panel A summarizes personal characteristics and residence information in terms of

[12]But this figure might be biased. Details will be discussed in next subsection.

[13]In all the rest of this paper, cohorts refers to NCEE-year cohorts 1977-2008.

[14]The college admission rate in each is a nationwide overall figure, thus restricting the sample to individuals in urban will substantially reduce the representativeness.

[15]Among cohorts later than 2008, there should be a certain amount of college students still in school.

[16]Most studies on return to education in China use log annual income as dependent variable due to lack of reliable measures of hourly wage rates. To make our estimation comparable to previous studies, all reported results in this paper are based on annual income and we do not present the estimates using wage rate as dependent variable.

[17]Years on current primary job.

provinces. Panel B includes four family background variables: parental education levels,[18] number of siblings and the Hukou status when taking the NCEE.[19] Panel C includes four NCEE-cohort specific variables matched to individuals through self-reported year of the last NCEE attempt. College admission rates and numbers of NCEE-takers are collected from Educational Statistic Yearbooks of China and China Statistical Yearbooks. The cohort-specific NCEE-age population is weighted average number of all potential NCEE-takers in each cohort.[20] Ratio of NCEE-takers is defined as the ratio of number of NCEE-takers to NCEE-age population. The trends of NCEE-age population, number and ratio of NCEE-takers are presented in figure 2.5. Overall, the ratio of NCEE-takers increases smoothly from less than 10 percent in 1983 to about 45 percent in 2008.

Sample Representativeness

Our sample is restricted to only include working individuals who once took the NCEE. If our sample is a random selection from the nationwide population, the cohort-specific share of college graduates should be close enough to the yearly college admission rate. But as figure 2.6 shows, the blue line representing our sample's share of college graduates is much higher than the green line depicting the official college admission rates for the population. Another important fact is that the trend of sample's college share is close to the population except that the increase in 1999 is not as remarkable as the population. To be specific, the cohort-specific college share differences between our sample and the population are quite stable at about 40 percent for cohorts before 1999, and about 30 percent for cohorts since 1999. This implies that the drastic expansion in 1999 has reduced the sample selection bias in cohort-specific college share.

To explore the causes of the college share differences between our sample and the population, we first check the cohort-specific college shares for all individuals who took the NCEE instead of only working individuals in our sample. As shown in figure 2.6, the full sample's cohort-specific college shares represented by red line are only lower than blue line very slightly. Thus, selection on working individuals is not a reason for the large sample bias in cohort-specific college shares.

Second potential cause is misreport on NCEE experiences. We check high school

[18]There are 246 and 322 missing values for mother's and father's education levels respectively. We assign missing values as 0 and set them as reference level in our analysis. Reported levels assigned 1 to 9 are never schooled, elementary , junior middle ,senior middle, vocational secondary, specialized secondary, polytechnic college(3-year college), undergraduate (4-year college) and graduate.

[19]A Hukou is a record in a government system of household registration required by law in China, and determines where citizens are allowed to live. Chinese citizens are broadly categorized as "rural" or "urban" Hukou. In this paper we assign 1 for urban Hukou and 0 for rural Hukou.

[20]It's computed based on the age-specific population and the distribution of ages when taking the NCEE. Age-specific population are collected from China Population Yearbook 2000 which reports the size of the population at each age by 1999. Distribution of NCEE-ages is based upon our sample as figure 2.4 shows. But this distribution is not applicable to early NCEE-cohorts 1977-1979 given that part of them took the NCEE at relatively older ages due to the 10 year suspension of NCEE during

and college graduates below age 50 who reportedly never took the NCEE.[21] Surprisingly, there are certain number of individuals that have completed high school or college education among them. As columns 2 and 5 show in table 2.2, there are 1301 college graduates and 2438 high school graduates reporting who never took the NCEE. China runs parallel off-campus higher education systems without requiring taking the NCEE, but CHIP does not differentiate between these two types of college degree. Therefore, it's plausible that most of the 1301 college graduates reporting no NCEE attempts could be off-campus college degree holders.[22] Normally, high school graduates take the NCEE to compete for college admission (Fan et al., 2010).[23] Thus, it's abnormal that the number of high school graduates reporting no NCEE attempts, 2438, is more than twice those reporting NCEE attempts, 1092. It's plausible that most of them are actually mis-reporters on NCEE attempts.

As we discuss above, there might be few misreports on NCEE attempts by college graduates while a lot of misreports by high school graduates. Therefore it's important to check whether those misreporting high school graduates are significantly different from those correctly reporting their NCEE attempts. Column 4 and 5 in table 2.2 show no evidence of significant differences between all variables. Thus, we believe that estimation of college premium with our sample will not be biased due to sample selection.

2.5 IV Estimation Based on 1999 Expansion

In this section, we use the discontinuous increase in college graduates share of cohort 1999 due to the drastic higher education expansion, arguing that individuals' unobserved characteristics are continuous at cohort 1999, to attribute the discontinuous increase in log annual earnings to the jump in college graduates share in cohort 1999. The estimated local average treatment effect (LATE) of college education for the group of individuals induced by the 1999 expansion carries strong policy implication. The estimated college premium, 75.7 percent and 4-year college premium, 92.1 percent (both with respect to high school graduates) demonstrate that the 1999 expansion substantially improve the earnings for those induced into college and provide a direct evidence of the effectiveness of the 1999 expansion on earnings improvement.

---

the Cultural Revolution 1966-1976. So we only compute NCEE-age population for cohorts from 1980.

[21]Checking those below age 50 is to limit the sample to those potential NCEE-takers later than 1977. Because some of those older than 50 may really never took the NCEE instead of misreport

[22]It's impossible to identify the exact number with limited information. But it's plausible that formal college degree holders have no incentive to misreport their NCEE experiences, so we believe most of the 1031 individuals are off-campus college degree holders.

[23]Fan et al. (2010) point out that most high school graduates take the NCEE but no data provided to support their statement. Here, I provide the data from China Statistical Year Books. The number of high school graduates is 2.52 million in 1998 and 2.63 million in 1999. The number of NCEE-takers is 3.2 million in 1998 and 2.88 million in 1999. The NCEE-takers are more than the high school graduates due to the re-takers. Even though we canâĂŹt tell the exact percentage of the NCEE-takers but itâĂŹs plausible that most high school graduates usually take the NCEE. This fact makes it not necessary to worry about the sample selection issue by using only NCEE-takers.

Methodology

Basically, our strategy is analogous to the regression discontinuity design which utilizes the fully (Sharp RD) or partly (Fuzzy RD) deterministic mechanism that a treatment is assigned to individuals once their observed characteristic (Forcing variable) pass an threshold (Eligibility cutoff). If individuals are unable to precisely manipulate the forcing variable, it will be reasonable to attribute the discontinuous jump in the outcome to the causal effect of the treatment (Lee and Lemieuxa, 2010). The inability of precise manipulation is equal to the assumption noted by Hahn et al. (2001), all omitted variables are continuous with respect to forcing variable. Heckman and Robb (1985) propose estimating the effect of the treatment by adding a flexible function of the forcing variable into the estimating equation in order to control for the potential omitted variables. Our case may be written as:

$$lnY_i = \alpha + \beta S_i + \gamma X_i + f(C_i) + \varepsilon_i \tag{2.1a}$$

$$Pr(S_i = 1|C_i \geq \bar{c}) > Pr(S_i = 1|C_i < \bar{c}) \tag{2.1b}$$

where lnY_i is the logarithm of annual earnings for individual i; S_i is a dummy variable indicating whether the individual has a college degree (or 4-year college degree); X_i is a vector of control variables including not only predetermined gender dummy, number of siblings, parental education levels and Hukou status when taking the NCEE but also tenure, squared tenure and province dummies; ε_i is the error term; C_i indicates the NCEE-cohort of individual i; $f(C_i)$ is 3-order polynomials; $\bar{c}$ is 1999 when China's college admission rate substantially increased. That passing the threshold $\bar{c}$ increases the probability of college entrance is modeled by equation (1b).[24] So our case is a fuzzy regression discontinuity design.

Fuzzy RD leads naturally to a simple 2SLS estimation strategy (Angrist and Pischke, 2008). The instrumental variable is the a dummy indicating individual's NCEE-year cohort is equal to or greater than the cutoff 1999. We call this IV as eligibility dummy in the rest of this paper. With assumptions of monotonicity and excludabilit, we're able to estimate an unbiased local average treatment effect (LATE).[25] And this LATE measures the causal effect of attending college for the group of individuals induced into college by the 1999 expansion.

Validity Tests

It's impossible to directly test the continuity assumption for all omitted variables with respect to forcing variable for a valid RD design since we have no observations for all omitted variables. Fortunately, two types of indirect tests are proposed. One

[24]Given that NCEE-cohort is highly correlated with age, so quadratic NCEE-cohort should fit the log earnings well. The first RD application to estimate educational returns by Oreopoulos (2006) just use a quadratic birth cohort. But in our case, cubic form fits college probability and log earnings better than quadratic form by goodness-of-fit statistics.

[25]By Imbens and Angrist (1994), monotonicity rules out the case where individuals were able to enter college before 1999 fail to be admitted after 1999. The excludability assumption demands that

is the test for continuous density of forcing variable proposed by McCrary (2008) and the other one is test for continuous predetermined covariates proposed by Lee and Lemieuxa (2010).

In our case the forcing variable is NCEE-cohort. So with our sample, the variation of its density reflects the expansion of the number of NCEE-takers in each cohort. Considering that cohort-specific averages of omitted variables should be more correlated with the cohort-specific ratio of NCEE-takers to NCEE-age population rather than the number of NCEE-takers, we test the continuity of the ratio at the cutoff cohort 1999. Another advantage of testing ratio continuity is that we can analyze the extent of the bias if there is a significant jump at cohort 1999. The cohort-specific ratio of NCEE-takers is shown in figure 2.5 in the data description section. As figure 2.5 shows, there is a slight decrease in the ratio at cohort 1999. We adopt a formal test following McCrary (2008) to examine whether the ratio of NCEE-takers shows any significant discontinuity at cohort 1999. The test is implemented by regressing the ratio on 3-order polynomials of NCEE-cohort and the eligibility dummy.[26] The estimated coefficient for eligibility is about -1.66 with t-statistics -1.60.[27] So we can infer that cohort-specific ratio of NCEE-takers shows no significant discontinuity at the cutoff cohort 1999.

We also follow Lee and Lemieuxa (2010) to estimate the seemingly unrelated regression for the four covariates (gender dummy, number of siblings, NCEE-Hukou status and tenure) to test whether they are jointly discontinuous at the cutoff cohort 1999. Tests are implemented for two samples separately. One sample includes college and high school graduates while the other sample includes only 4-year college and high school graduates. Results of the test are presented in table 2.3. All the estimated coefficients of eligibility dummy are not statistically different from zero. By the χ^2 tests, we can not reject the null hypothesis that coefficients for eligibility dummy are jointly equal to zero.

Since none of the two types of test shows evidence of significant discontinuity at the cutoff cohort 1999, our estimates identify unbiased LATE of college education for the group induced into college by the 1999 expansion.

Results: Graphical Presentation and Regression Analysis

Before discussing formal regression results, we first present the two sets of graphs to show that the treatment variables (college or 4-year college) and outcome variable (log annual earnings) are discontinuous at the cutoff cohort 1999. Figure 2.7 is for the sample including all college and high school graduates while figure 2.8 is for the sample including only 4-year college and high school graduates. Dots in the left graphs

the eligibility dummy can only affect the dependent variable through its impact on the treatment.

[26] But we only use the ratio of cohorts later than 1981 because the ratio for early cohorts might be computed with substantial bias. That's because, as we discussed before, potential NCEE-age population for early cohorts might be substantially biased due to the existence of much older NCEE-takers due to the 10 years of suspension of the NCEE during the Culture Revolution.

[27] Table of the regression result is not presented here.

are cohort-specific shares of college and 4-year college graduates respectively. In the right graphs, dots represent the cohort-specific average log annual earnings. All fitted lines are obtained from the regression of corresponding dependent variables(college, 4-year college or log annual earnings) on a 3-order polynomials of NCEE-cohort and eligibility dummy.

Both figures show very clear discontinuities of the outcome (log annual earnings) and treatment variables(college or 4-year college dummy) at the cutoff cohort 1999. Even if the inter-cohort trend of log annual earnings is not solely affected by cohort-specific composition of education levels, we can still see some evidence of the effect of the educational composition.[28] As the cohort-specific shares of college graduates increased rapidly from the earliest cohort 1977 to 1985, the corresponding average log annual earnings also increase substantially. For the stable cohorts 1986 to 1998, earnings decrease slightly due to the decreasing average tenure/age. After the cutoff cohort 1999, the shares of college graduates are shifted up by the 1999 expansion, the log annual earnings are also shifted up as expected by the compositional improvement in education levels. The rapid decrease of the earnings should be attributed to decreasing average tenure/age and average ability.[29]

The identification of the LATE is just attributing the discontinuity of log annual earnings at the cutoff to the jump of college share at the cutoff. And the computation of the magnitude is dividing the jump of log annual earnings by the jump of college share, which may be implemented by 2SLS estimation with eligibility dummy as instrumental variable for the treatment variables (college or 4-year college dummy).

In table 2.4, we present the results of the first stage regressions in columns 1-4 and results of reduced form regressions in columns 5-8. Specifications for column 1,3,5 and 7 only control for province dummies while full covariates including gender, quadratic tenure, number of siblings, parental education levels, NCEE-Hukou status and province dummies are controlled for in columns 2,4,6 and 8. All estimated coefficients of our instrumental, eligibility dummy, are all very significantly positive from 0.106 to 0.148 in the first stage estimations. The inclusion of full covariates reduces the effects and F-statistics of the instrumental variable slightly. F-statistics are all larger than 10, which means we have no issue of weak IV in our case according to (Angrist and Pischke, 2008). Estimated coefficients of our IV in reduced form regressions are all positive and statistically significant at 5 to 10 percent significance level.

Our estimates of college and 4-year college premium are presented in columns 5-8 in table 2.5. With respect to high school graduates, ceteris paribus, college graduates receive 0.757 additional log annual earnings on average, while 4-year college graduates earn 0.921 more.[30] Inclusion of the full covariates increases the estimates slightly to

[28]NCEE-cohort is correlated to age and tenure, thus the trend is also partially determined by the age-earnings or tenure-earning profiles.

[29]As figure 2.5 shows, the ratio of NCEE-takers increase rapidly since 1999, which will definitely lower the cohort-specific average ability.

[30]This figure indicates 113 percent additional annual earnings in terms of the computation $e^{0.757} - 1$. But following the literature we just report the coefficients directly estimated and make comparison

0.828 and 0.963. Even if these estimates are only statistically significant at 5 to 10 percent confidence level, they are all very economically significant. Especially when compared with their corresponding OLS estimates presented in columns 1-4, we find that our estimates are more or less doubled. Our estimates are consistent with the existing studies that usually find IV estimates are less precise and larger than the OLS estimates (Card, 1999b; Heckman et al., 2006). What's more interesting is that very close results are revealed by the only one existing RD-IV study of returns to higher education in China by Fan et al. (2010). Utilizing a RD design based on China's score-based college admission mechanism and a different data set, they obtain OLS estimate and RD-IV estimate of 4-year college premium with respect to high school graduates, 0.52 and 1.12, for male. Their estimate of the LATE, which measures the returns to college education for the group whose NCEE-scores are at the cutoff threshold, provides some insight into the effect of the higher education expansion because the group at the cutoff are most likely affected by the expansion. Our estimate of the LATE measures the returns for a more specific group that are induced into college by the drastic 1999 expansion. Therefore, our LATE estimate shed light to the evaluation of the effect of 1999 expansion more directly. Based on our result, 0.757 for college graduates and 0.921 for 4-year college graduates, the 1999 expansion has indeed substantially benefited the group of individuals induced into college.

2.6 IV Estimation Based on Admission Rates

In last section, we only utilize the discontinuity of college admission rate at the cutoff cohort 1999 to estimate a very specific LATE for the group of individuals induced into college by the 1999 expansion. The estimated LATE is very policy relevant but less representative for the population. From figure 2.6, it's easy to notice that the cohort-specific college admission rate varies between a large range, less than 5 percent in 1977 and slightly over 60 percent in 2003. It's implied that more and more individuals have been induced into college from early to recent NCEE-cohorts and the proportion being affected is really large. Even if we look at the blue line representing the sample used in this paper, we can still see a substantial variation of the cohort-specific share of college graduates from about 40 percent to about 90 percent. Therefore, using cohort-specific college admission rate as instrumental variable for education will result in a more population-representative estimate if it is a valid IV. In this section, we discuss the validity of college admission rate as IV for education variable and construct an additional interaction IV following the a series of studies in earnings effects of education (Card and Krueger, 1992; Lemieux and Card, 2001; DuFLo, 2001; Maurin and McNally, 2008). Earnings return to an additional year of higher education as well as college premium will be estimated by using two types of measure for education, self-reported years of formal education and college dummy.

with existing estimates for convenience.

Methodology

Consider an observed covariate Z_i that affects schooling but has no direct effect on earnings. The IV estimate of the return to schooling can be obtained by 2SLS method. First, run OLS regression by equation (2a) and obtain the predicted $\hat{S}_i$. Then, substitute S_i in equation (2b) with $\hat{S}_i$. The OLS estimate for β is just the IV estimate we are seeking. For binary endogenous S_i, college dummy in this paper, we apply method of propensity score IV well discussed by Wooldridge (2010). We first generate predicted probabilities of college attendance from the probit regression on equation (2a) and then use variable of the predicted probability as instrumental variable for college dummy.

$$S_i = \alpha_s + X_i\gamma_s + Z_i\rho_s + \eta_i \tag{2.2a}$$
$$lnY_i = \alpha_y + X_i\gamma_y + \beta S_i + \varepsilon_i \tag{2.2b}$$

IV Candidates

The basic instrumental variable used is cohort-specific college admission rate as presented by the green line in figure 2.6. By its definition, ratio of the number of college enrollment to the number of NCEE-takers in each cohort, the variation comes from two sources: variation of the number of college enrollment and variation of the number of NCEE-takers. Given that the scale of the college enrollment is administered by central government in China, its variation reflects the governmental plans on higher education. To be specific, the scale of college enrollment increased mildly until 1998 and rapidly since 1999 when the ambitious expansion launched. The number of NCEE takers is partially determined by the exogenous demographic trend of the NCEE-age population, and partially determined by the capacity of high schools which is administered by the local governments. So, the variation in the number of cohort-specific NCEE-takers is also exogenously determined.

Intuitively, in a year with higher college admission rate, higher share of the NCEE-takers will be admitted. As reflected in a cross section data set like the sample in our analysis, the share of college graduates in the cohort with higher admission rate should be higher. This expectation is well depicted in figure 2.6 that the inter-cohort trends of blue and green lines are very close. This straightforward positive correlation meets the first condition of valid IV that an IV must affect the variable instrumented. Formal test will be presented and discussed in next subsection. But if there exists cohort fixed earnings effects correlated with cohort-specific college admission rate, the validity of it as an IV will be ruined. As we know, the college admission rate is a cohort-specific variable which has no variation across individuals within each cohort. Thus, it's impossible for us to control for the cohort fixed effects directly in the earnings specification.

Our first solution for this issue is to control for a 3-order polynomials of NCEE-cohort, which is based on an assumption that the cohort fixed effects are smooth across cohorts. Considering that cohort-specific average unobservables, e.g ability, are likely correlated with the cohort-specific ratio of NCEE-takers, the assumption seems plausible by looking into figure 2.5 where the inter-cohort trend of the ratio of NCEE-takers is quite smooth and follows a low order polynomial form approximately.

We also use interaction between NCEE-Hukou status and college admission rate as another instrumental variable by which we can relax the assumption of smooth cohort fixed effects and control for cohort dummies directly.[31] The validity of this interaction term as IV relies upon the assumption that the differences in cohort fixed earnings effects between two NCEE-hukou groups (urban and rural Hukou) are constant across cohorts or uncorrelated to college admission rates at least. We use a set of figures 2.9(a)-2.9(c) to interpret the identification strategy underlying this interaction IV.

In figure 2.9(a), square and triangle dots represent cohort-specific shares of college graduates for group with urban NCEE-Hukou and group with rural NCEE-Hukou respectively. Green line and orange line are their corresponding quadratically fitted lines. It's quite clear that the differences in shares of college graduates between the two NCEE-Hukou groups decreases as college admission rate increases. As figure 2.9(b) shows, similar patterns exist for another measure of education, the self-reported years of formal education. Considering that the average starting education level for the group with rural NCEE-hukou is much lower than the urban group, it's plausible to argue that the convergence just reflects a natural catch-up process instead of narrowing differences in cohort-specific fixed effects between the two groups.

AS figure 2.9(c) shows, the earnings gap between the urban and rural NCEE-Hukou groups decreases as college admission rate increases. It's natural to link the narrowing earnings gaps to the narrowing educational gaps shown in figure 2.9(a) and 2.9(b). But to identify the effect of education, we must be careful about a fact that there might be other factors also playing certain roles in the earnings convergence process. Thus, when using interaction IV, we must control for interactions between NCEE-hukou status and other regressors, especially the experience terms.

With two IV candidates, we can implement our estimations separately first and then use both to examine the over-identification by which to double check the validity of our IVs.

Estimation Results

As we discussed in last subsection, with different instrumental variables used we include different control variables in addition to the basic set of covariates (gender, quadratic tenure, number of siblings, parental education levels, NCEE-hukou status, urban dummy and province dummies). Specifically, we control for 3-order polynomials of NCEE-cohort when using college admission rate as IV, control for interactions between NCEE-hukou status and quadratic tenure as well as cohort fixed effect when using interaction IV, and control for interactions between NCEE-hukou status and quadratic tenure as well as 3-order polynomials of cohort when using both IVs.

Before discussing the IV estimates we first examine the results from the first stage and reduced form estimations. They are presented in table 2.6 where columns 1-3

[31] Hukou refers the residence status, urban and rural. NCEE-Hukou is the information surveyed

and 4-5 are first stage estimations with years of education and college dummy as endogenous variable respectively, and columns 7-9 are reduced form estimations. In columns 1,4 and 7 we present results when using college admission rate as IV. The effects of college admission rate on years of education and college attendance are both quite statistically significant even if the smooth cohort fixed effects are controlled for by a 3-order polynomials. The F-statistics and Likelihood ratio statistics are both larger than 10, which means college admission rate can be rejected to be a weak IV. In the reduced form regression in column 7, college admission rate positively affects log annual earnings but not quite statistically significant. In columns 2,5 and 8 we present results using interaction IV. Consistent with our previous graphical analysis ,in columns 2,5 and 8, the negative estimated coefficients for this interaction IV imply that education premiums as well as the log earnings premiums for the group with urban NCEE-hukou decreases as college admission rate increases. We also notice that the F and Likelihood ratio statistics of this interaction IV, are both larger than those of college admission rate. In columns 3,6 and 9 we present results for using both IVs together. Estimated coefficients for both Ivs are very statistically significant, 0.032, -0.028 for years of education in column 3 and 0.023, -0.019 for college attendance in columns 6. These estimates also imply that the overall increase of the education along with the increase of college admission rate are mostly attributed to the education improvement of the group with rural NCEE-hukou. In the reduced form estimation as the last column shows, the estimated effects of the two IVs are also consistent with our expectation that college admission rate positively affect earnings while the earnings premiums for the group with urban NCEE-hukou decrease along with the increase of college admission rate.

Our IV estimates of the returns to one additional year of higher education are presented in columns 4-6 in table 2.7. Corresponding OLS estimates are also presented in columns 1-3. The OLS estimates of the return to one year of higher education are very precise at about 7 percent and are unaffected by the inclusion of different sets of additional control variables. To be specific, the specifications for columns 1 and 4 control for 3-order polynomials of cohort. Fixed NCEE-cohort effects and interactions between NCEE-hukou status and quadratic tenures are controlled for in columns 2 and 5. Both 3-order polynomials of cohort and interactions between NCEE-hukou status and quadratic tenures are controlled for in columns 3 and 6.

All of the three IV estimates are more than doubled but less precise relative to OLS estimates as column 4-6 show. This is consistent with our first set of IV estimates based on 1999 expansion as well as the studies in the literature that compare OLS and IV estimates. The most meaningful finding is that the two IV estimates,0.175 and 0.15, in columns 4 and 5 with different IVs (college admission rate as IV in column 4, interaction between NCEE-hukou and college admission rate as IV in column 5) are very close. This substantially increases our confidence in the validity of the two IVs used. When using both IVs in column 6, the estimated returns is much more precise at 16.1 per cent. And formally, the over-identification test statistics is so small that

in CHIP 2013 about the residence status when one took the NCEE.

we can not reject the null hypothesis that our IVs are valid.

Estimates of college premium are presented in table 2.8. Similarly, OLS estimates and IV estimates are presented separately in columns 1-3 and 4-6. Specifications are same as those used in table 2.7 except for the measure of the education variable of our interest. Following the method well discussed in Wooldridge (2010), we use predicted college attendance probability from the first stage probit regression as instrumental variable for college dummy.

It's revealed again that the OLS estimates of college premium are quite precise at about 30 per cent and unaffected by different specifications. Considering that in this paper college dummy is defined as 1 for both 3-year and 4-year college graduates and as 0 for high school graduates, the average difference in the years of formal education between college and high school graduates should be larger than 3 but smaller than 4[32]. However, OLS estimates of college premium, 30 percent are more than 4 times larger than OLS estimates of returns to one additional year, 7 per cent. This inconsistence reflects somewhat bias even if it's not very clear how and to what extent these estimates are biased. Interestingly, this inconsistence disappears among those IV estimates. Comparing IV estimates of college premium 0.645, 0.506 and 0.575 with IV estimates of returns to additional year of higher education 0.175, 0.15 and 0.161, we find that estimated college premiums are all about 3.6 times larger than estimated returns to one additional year of higher education. We believe this fact serves as an interesting indirect proof for the validity of our instrumental variables.

2.7 Conclusion

In this paper, we utilize the unique long-lasting higher education expansion since 1977 in China to identify the causal effects of higher eduction on earnings of Chinese. The central government-controlled college enrollment scale and demographically affected college applicants scale result in a plausible exogenous, or predetermined at least, variation of cohort-specific college admission rates. The drastic increase of the college admission rate in 1999 not only demonstrates the governmental control on higher education but also provides a rare opportunity for us to identify a more specific local treatment effect of college education on earnings. Based on these facts of the higher education expansion, we implement the IV estimation in two ways: 1999 expansion and yearly admission rate.

China Household Income Project 2013 surveys a series NCEE-related questions, which make it possible for us to match the cohort-specific college admission rates to each individual accurately. First, we use NCEE-cohort as forcing variable and an indicator of being later than cohort 1998 as eligibility dummy. Applying 2SLS estimation procedure, we estimate the LATEs of college education at 75.7 percent and, more specifically, 4-year college education at 92.1 percent. Considering that these estimates are actually the LATEs of college education for the specific group of individuals induced into college due to the drastic higher education expansion in

[32]In our sample, high school graduates' average years of education is 11.4 while college graduates'

1999, these large estimated effects have very strong policy implication that the 1999 expansion has indeed substantially benefited those compliers on earnings. But this is also a limitation for us to make inferences more generally.

Considering that the IV estimates are a weighted average of the effects for individuals who are affected by the instruments, our conventional IV estimates reduce the limitation substantially. That's because the two instruments used, cohort-specific college admission rate and the interaction between the admission rate and NCEE-Hukou status, affect individuals from all cohorts instead of only those of cohort 1999. To deal with the potential issue of correlated cohort fixed effects, we first assume and control for a smooth inter-cohort trend when only using college admission rate as instrument, and then relax the assumption to control for unrestricted cohort fixed effects in an alternative specification with the interaction term as instrument. Both strategies result in quite close estimates. Last, when using both IVs, we obtain more precise estimates and pass the over-identification test. To be specific, returns to one additional year of higher education is estimated at 16.1 per cent and college premium is estimated at 57.5 per cent. That our IV estimates are larger than corresponding OLS estimates are consistent with the literature that finds IV estimates of educational returns are usually larger than OLS estimates. One explanation emphasizes the potential measurement error in the education variable that attenuates the OLS estimates. The other one takes into account the heterogeneity in the return to education and attributes the higher IV estimates to the potential higher returns for those affected by instrumental variables.

Comparing with existing studies, our estimates are much more policy relevant in China where the higher education experienced mild expansion form 1977 to 1998 and drastic expansion since 1999. Our IV estimates offer strong evidences on the earnings effect on individuals affected by the higher education expansion. Thus, by our findings, the long-lasting expansion has indeed substantially improved the earnings of those who obtained college education due to the expansion.

average is 16. The specific difference is 3.6 years.

Figures

Figure 2.1: Changes in China's wage determination 1988-2009

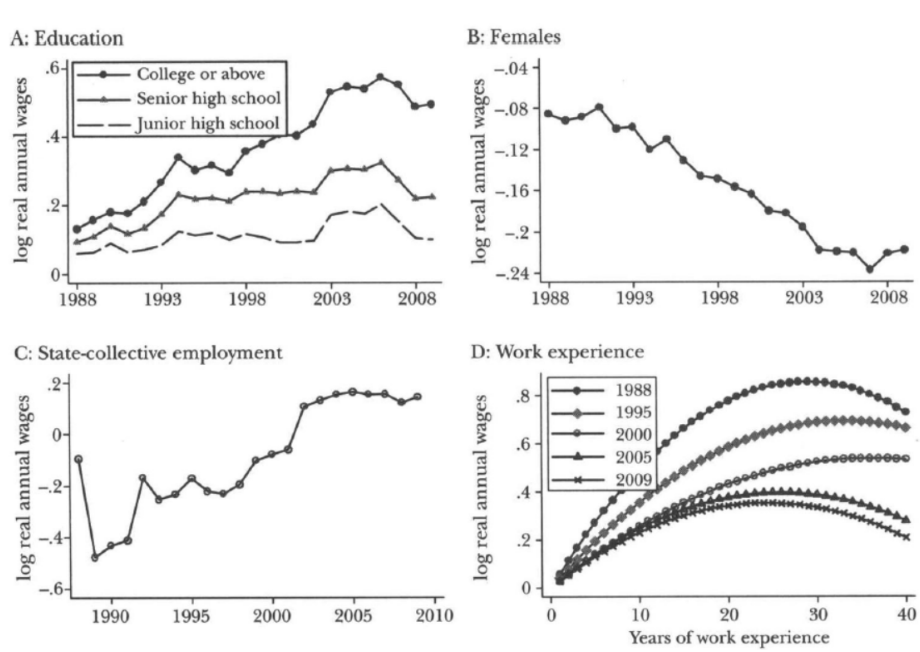


Figure 2.2: China's Higher Education Expansion:Scales

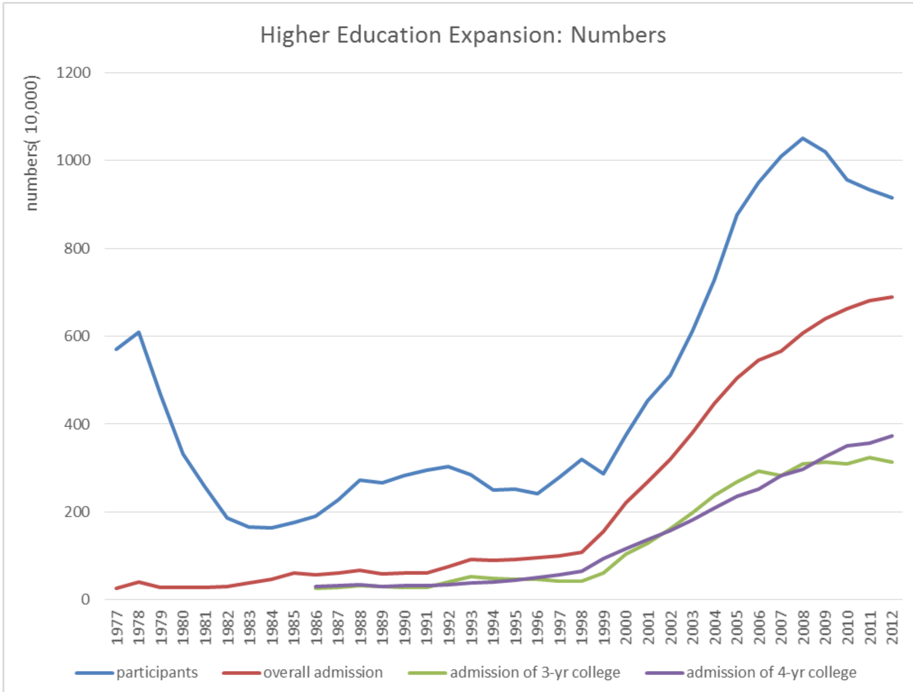


Figure 2.3: China's Higher Education Expansion:Rates

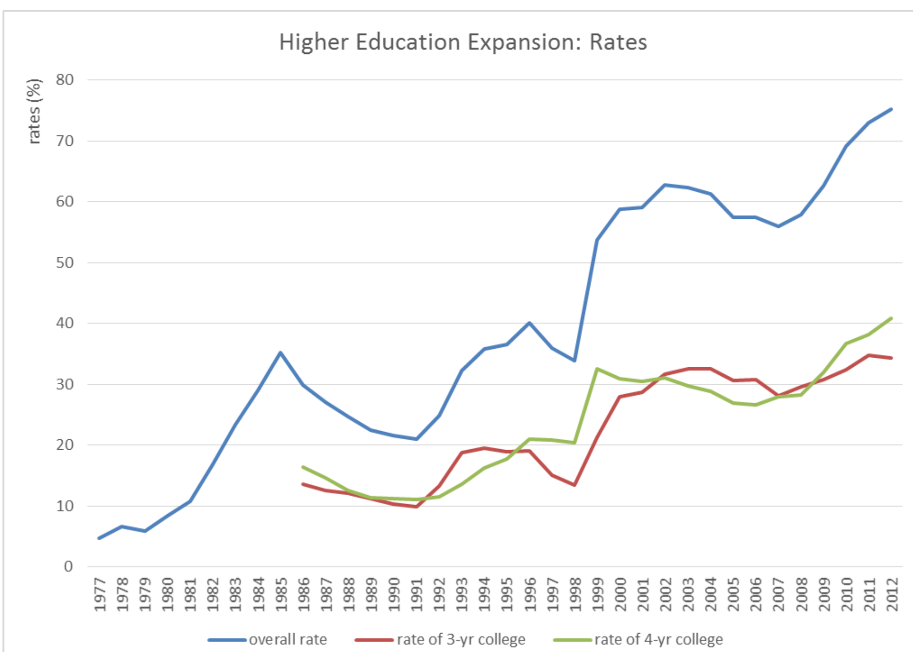


Figure 2.4: Distribution of the Age When Taking the NCEE

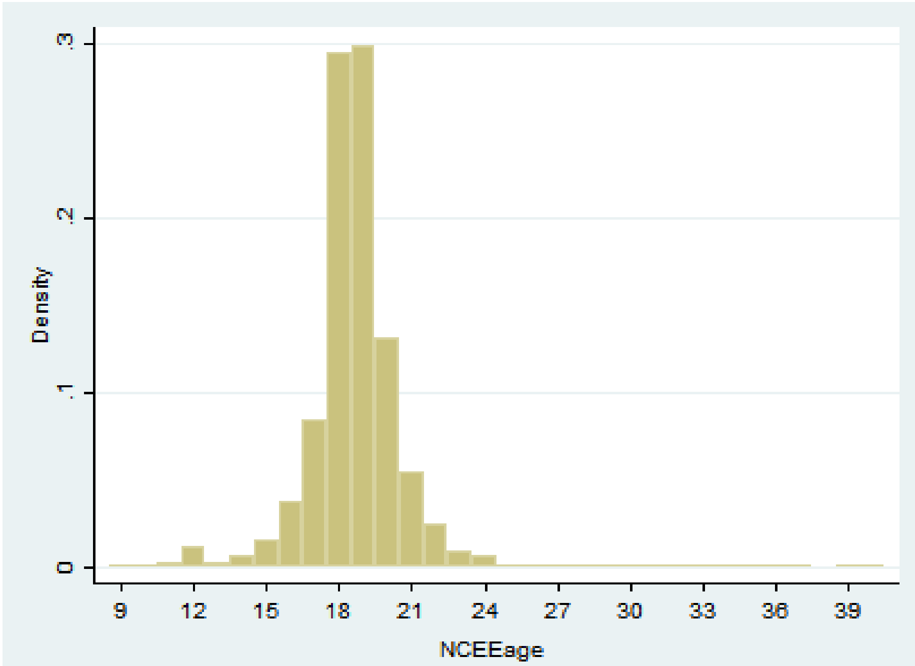


Figure 2.5: Cohort-Specific NCEE-Age Population, Number and Ratio of NCEE-Takers

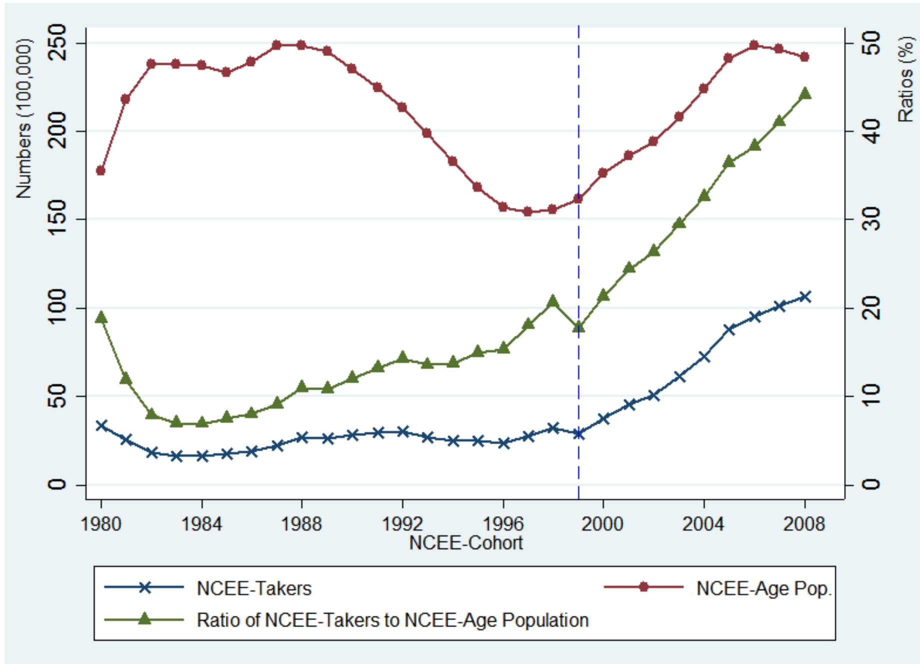


Figure 2.6: Cohort-Specific Shares of College Graduates of Sample and Population

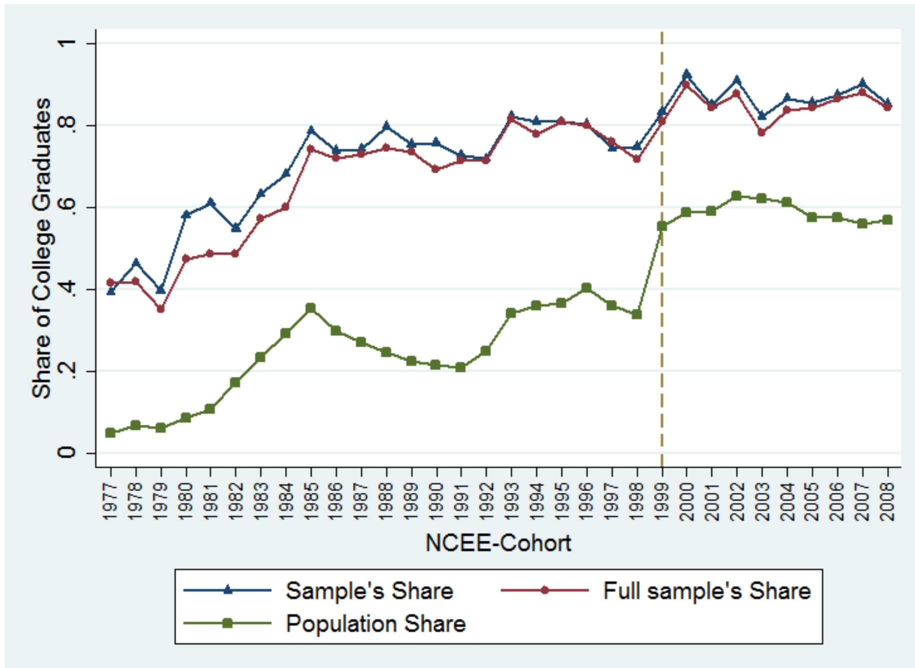


Figure 2.7: Test of the discontinuity of college attendance and log annual earnings at cutoff cohort 1999

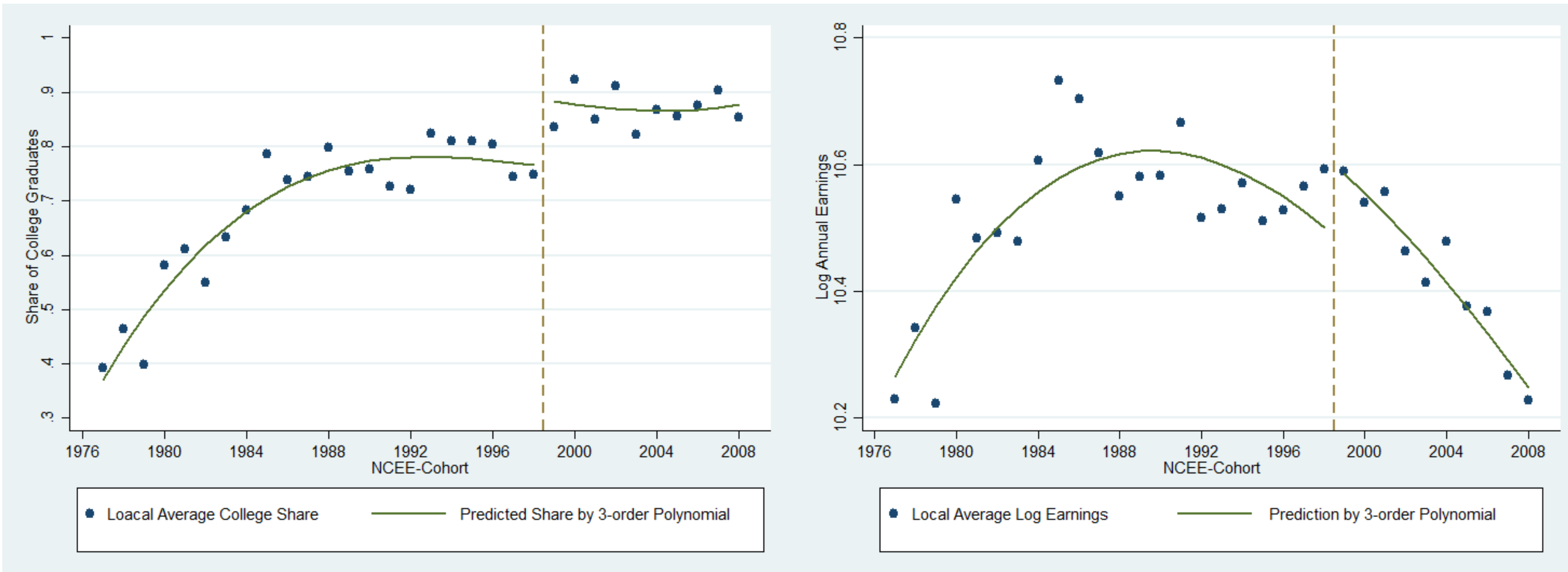


Figure 2.8: Test of the discontinuity of 4-year college attendance and log annual earnings at cutoff cohort 1999

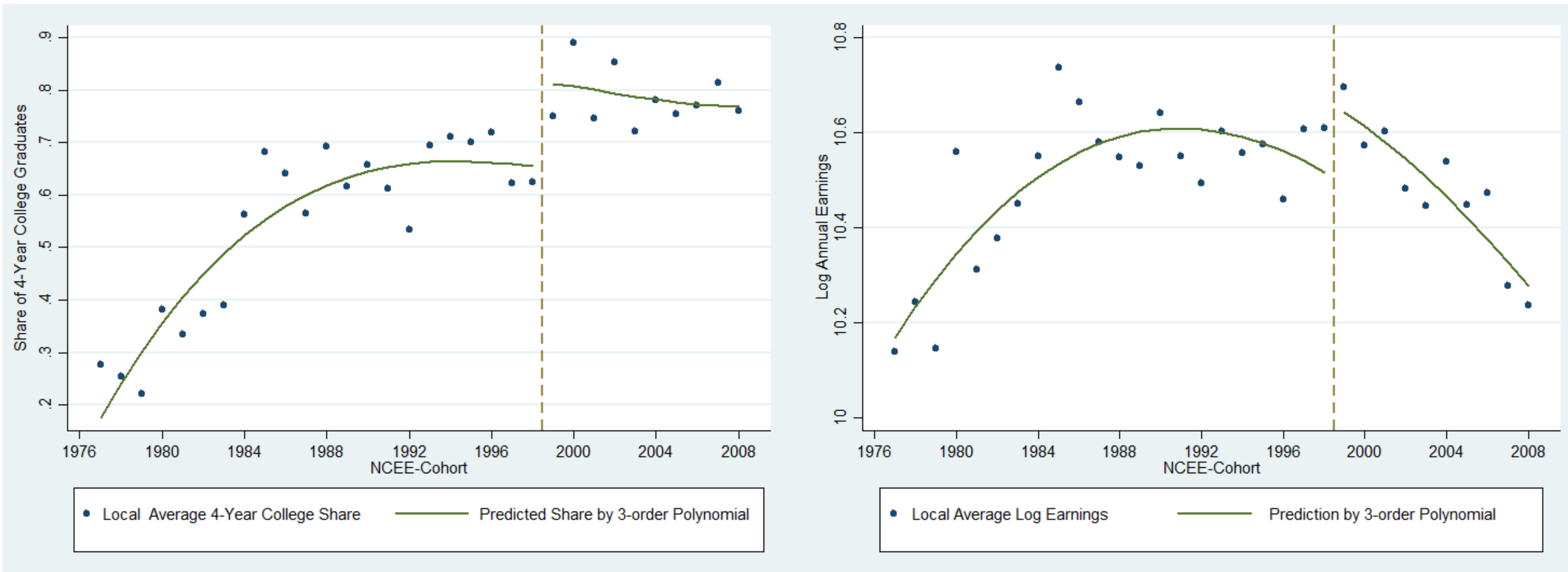
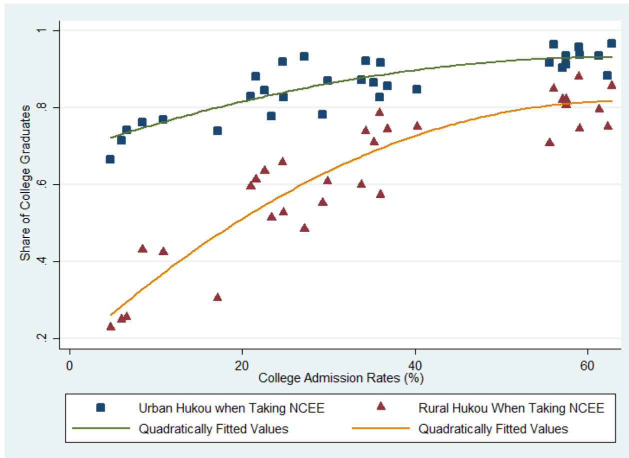
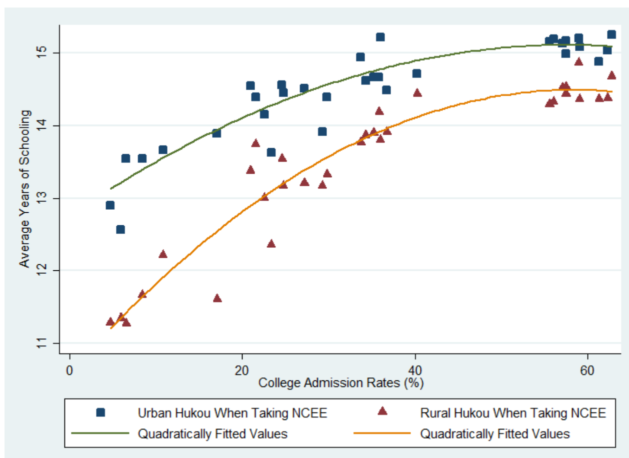


Figure 2.9: NCEEHukou-specific Local Averages of Education and Earnings

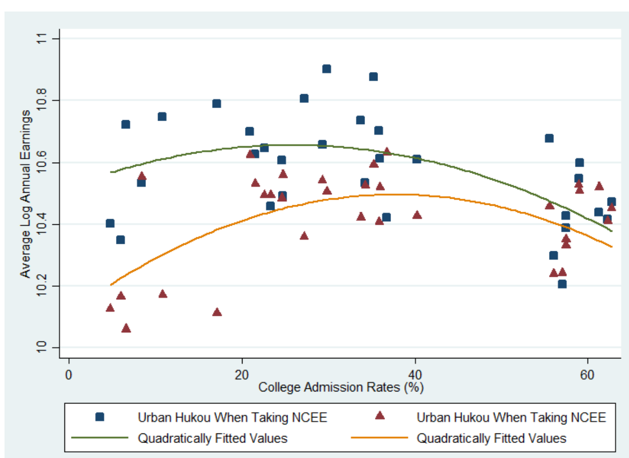
(a) NCEEHukou-specific trends of local share of college graduates



(b) NCEEHukou-specific trends of local average years of education



(c) NCEEHukou-specific trends of local average log annual earnings



Tables

Table 2.1: Summary Statistics

Panel A:Personal Characteristics	1977-1987 (1)	1988-1998 (2)	1999-2008 (3)	Urban (4)	Rural (5)	Total (6)
Average Anual Income (Yuan)	36,315	38,561	32,860	38,949	28,001	35,242
Average Wage Rates	17.64	17.81	14.88	17.81	13.46	16.28
Years of Education	12.97	14.22	14.79	14.53	13.46	14.21
	(2.29)	(1.84)	(1.59)	(1.79)	(2.17)	(1.97)
College(3- and 4-Year)	0.60	0.77	0.87	0.86	0.59	0.78
	(0.49)	(0.42)	(0.33)	(0.35)	(0.49)	(0.41)
4-Year College	0.29	0.43	0.47	0.48	0.27	0.42
	(0.45)	(0.50)	(0.50)	(0.50)	(0.44)	(0.49)
Age	49.20	38.60	27.62	36.90	33.24	35.81
	(3.77)	(4.17)	(3.32)	(8.89)	(9.92)	(9.36)
Tenure	18.82	12.92	4.90	12.09	6.67	10.47
	(11.34)	(6.84)	(3.37)	(9.27)	(6.82)	(8.96)
Male	0.69	0.58	0.53	0.56	0.64	0.59
	(0.46)	(0.49)	(0.50)	(0.50)	(0.48)	(0.49)
Urban	0.74	0.82	0.60			0.70
	(0.44)	(0.38)	(0.49)			(0.46)
Province Dummies:						
Beijing	0.12	0.11	0.10	0.13	0.06	0.11
Shanxi	0.08	0.05	0.07	0.07	0.04	0.06
Liaoning	0.04	0.04	0.03	0.04	0.01	0.04
Jiangsu	0.11	0.12	0.12	0.10	0.16	0.12
Anhui	0.06	0.04	0.04	0.04	0.05	0.05
Shandong	0.07	0.08	0.08	0.08	0.08	0.08
Henan	0.08	0.10	0.09	0.09	0.09	0.09
Hubei	0.09	0.09	0.07	0.07	0.10	0.08
Hunan	0.08	0.06	0.07	0.06	0.09	0.07
Guangdong	0.06	0.09	0.13	0.09	0.14	0.10
Chongqing	0.06	0.06	0.06	0.07	0.03	0.06
Sichuan	0.03	0.05	0.06	0.05	0.04	0.05
Yunnan	0.07	0.05	0.05	0.05	0.06	0.05
Gansu	0.05	0.05	0.05	0.04	0.05	0.05
Panel B: Family Background Variables						
Father's Education Levels	2.66	2.87	3.16	3.22	2.35	2.96
	(1.87)	(1.83)	(1.88)	(2.00)	(1.36)	(1.87)
Mother's Education Levels	2.13	2.41	2.85	2.78	2.03	2.55
	(1.56)	(1.54)	(1.64)	(1.72)	(1.22)	(1.62)
No. of Siblings	2.86	1.69	0.88	1.50	1.73	1.57
	(1.59)	(1.41)	(0.95)	(1.47)	(1.52)	(1.48)
NCEE-Hukou	0.48	0.55	0.49	0.70	0.06	0.50
	(0.50)	(0.50)	(0.50)	(0.46)	(0.24)	(0.50)
Panel C: Cohort-Specific Variables						
College Admission Rate(%)	17.64	30.51	58.43	38.91	45.07	40.76
	(10.60)	(6.80)	(2.28)	(18.14)	(18.41)	(18.44)
Number of NCEE-takers(100,000)	30.73	27.63	75.25	44.83	64.38	50.69
	(15.68)	(2.43)	(26.61)	(27.40)	(32.63)	(30.41)
NCEE-age Population(100,000)	210.60	195.34	219.76	207.45	216.68	210.21
	(36.93)	(35.72)	(29.75)	(34.55)	(34.98)	(34.93)
Ratio of NCEE-takers(%)	15.57	14.61	33.16	21.37	29.01	23.66
	(11.35)	(2.97)	(8.36)	(11.21)	(12.32)	(12.07)
Observations	830	1145	1749	2608	1116	3724

Table 2.2: Summary Statistics for the Sample Including all College and High School Graduates

	College Graduates			High School Graduates		
	NCEE=1 (1)	NCEE=0 (2)	Total (3)	NCEE=1 (4)	NCEE=0 (5)	Total (6)
Being Employed	0.71	0.88	0.75	0.71	0.76	0.75
	(0.45)	(0.32)	(0.43)	(0.45)	(0.43)	(0.43)
Average Annual Income (Yuan)	35,242	34,892	35,134	25,084	26,108	25,848
Age	30.20	36.07	31.56	34.99	36.92	36.32
	(8.56)	(8.36)	(8.86)	(10.02)	(9.45)	(9.67)
Male	0.52	0.50	0.52	0.54	0.53	0.53
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Urban	0.66	0.81	0.69	0.40	0.51	0.48
	(0.47)	(0.39)	(0.46)	(0.49)	(0.50)	(0.50)
Father's Education	3.28	2.98	3.21	2.35	2.26	2.29
	(1.96)	(1.82)	(1.93)	(1.50)	(1.40)	(1.43)
Mother's Education	2.89	2.52	2.80	2.02	1.91	1.95
	(1.73)	(1.57)	(1.70)	(1.27)	(1.17)	(1.20)
No. of Siblings	1.18	1.56	1.27	1.92	1.95	1.94
	(1.23)	(1.44)	(1.29)	(1.52)	(1.49)	(1.50)
Observations	4325	1301	5626	1092	2438	3530

Table 2.3: Validity Test for Joint Significance of Baseline Covariates

	Sample of college and high school				Sample of 4-year college and high school			
	Male	No. of Siblings	NCEE-Hukou	Tenure	Male	No. of Siblings	NCEE-Hukou	Tenure
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
cohort	-0.026**	-0.031	0.028***	0.130	-0.019	-0.036	0.028**	0.430**
	(0.010)	(0.025)	(0.010)	(0.140)	(0.012)	(0.030)	(0.012)	(0.170)
$cohort^2$	0.100	-0.654***	-0.114	-4.407***	0.054	-0.653***	-0.093	-5.525***
	(0.074)	(0.181)	(0.074)	(1.014)	(0.088)	(0.220)	(0.088)	(1.245)
$cohort^3$	-0.013	0.168***	0.007	0.731***	-0.004	0.167***	0.003	0.850***
	(0.014)	(0.035)	(0.014)	(0.196)	(0.017)	(0.043)	(0.017)	(0.243)
Eligibility	-0.027	-0.104	0.060	-0.234	-0.022	-0.046	0.069	0.119
	(0.038)	(0.093)	(0.038)	(0.520)	(0.046)	(0.115)	(0.046)	(0.654)
Observations	3,724	3,724	3,724	3,724	2,378	2,378	2,378	2,378
R-squared	0.026	0.352	0.053	0.445	0.030	0.371	0.059	0.389
χ^2-test	$\chi^2 = 3.82,\ Prob. > \chi^2 = 0.43$				$\chi^2 = 2.35,\ Prob. > \chi^2 = 0.67$			

Notes:*** p<0.01, ** p<0.05, * p<0.1 Robust standard errors in parentheses. All specifications include province dummies. Estimates for their effects are not presented in this table.

Table 2.4: First Stage and Reduce Form Estimations

	First Stage Estimation				Reduced Form Estimation			
	College	College	4-year College	4-year College	Log Earnings	Log Earnings	Log Earnings	Log Earnings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eligibility	0.120***	0.106***	0.148***	0.129***	0.091*	0.088*	0.137**	0.124**
	(0.030)	(0.029)	(0.041)	(0.038)	(0.053)	(0.051)	(0.063)	(0.060)
cohort	0.066***	0.057***	0.072***	0.052***	0.063***	0.050***	0.072***	0.049**
	(0.009)	(0.008)	(0.011)	(0.010)	(0.017)	(0.016)	(0.021)	(0.019)
$cohort^2$	-0.334***	-0.259***	-0.336***	-0.237***	-0.332***	-0.229**	-0.337**	-0.220
	(0.067)	(0.060)	(0.082)	(0.073)	(0.117)	(0.109)	(0.146)	(0.134)
$cohort^3$	0.053***	0.043***	0.049***	0.039***	0.038*	0.026	0.033	0.025
	(0.013)	(0.011)	(0.016)	(0.014)	(0.022)	(0.021)	(0.027)	(0.025)
F-statistics	15.936	13.939	12.735	11.462				
Full Covariates		Yes		Yes		Yes		Yes
Observations	3,724	3,724	2,378	2,378	3,724	3,724	2,378	2,378
R-squared	0.115	0.236	0.171	0.320	0.092	0.181	0.090	0.201

Notes:*** p<0.01, ** p<0.05, * p<0.1 Robust standard errors in parentheses. All specifications control for province dummies. Full covariates include male dummy, quadratical tenure, NCEE-Hukou, number of siblings, dummies indicating father's and mother's education levels. Estimates for their effects are not presented in this table. "Yes" indicates the full controls are included.

Table 2.5: OLS and RD-IV Estimates of College and 4-year College Premiums

	(OLS estimation) Log Annual Earnings				(RD-IV estimation) Log Annual Earnings			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
College	0.448***	0.342***			0.757*	0.828*		
	(0.030)	(0.031)			(0.445)	(0.500)		
4-year College			0.525***	0.408***			0.921**	0.963*
			(0.032)	(0.036)			(0.436)	(0.497)
cohort	0.030**	0.025*	0.029	0.020	0.013	0.003	0.006	-0.002
	(0.014)	(0.014)	(0.018)	(0.017)	(0.027)	(0.026)	(0.029)	(0.025)
$cohort^2$	-0.151	-0.096	-0.114	-0.067	-0.079	-0.014	-0.028	0.008
	(0.097)	(0.093)	(0.121)	(0.116)	(0.137)	(0.124)	(0.148)	(0.133)
$cohort^3$	0.009	0.005	0.000	-0.000	-0.002	-0.009	-0.012	-0.013
	(0.019)	(0.018)	(0.024)	(0.023)	(0.024)	(0.023)	(0.027)	(0.026)
Male		0.204***		0.201***		0.210***		0.201***
		(0.021)		(0.027)		(0.023)		(0.028)
Tenure		0.027***		0.030***		0.020**		0.018
		(0.005)		(0.007)		(0.009)		(0.013)
$Tenure^2$		-0.037**		-0.045**		-0.036**		-0.031
		(0.015)		(0.021)		(0.015)		(0.025)
NCEE-Hukou		-0.013		-0.032		-0.071		-0.106
		(0.024)		(0.031)		(0.064)		(0.074)
No. of Siblings		-0.034***		-0.048***		-0.027**		-0.038**
		(0.009)		(0.011)		(0.011)		(0.015)
Observations	3,724	3,724	2,378	2,378	3,724	3,724	2,378	2,378
R-squared	0.151	0.211	0.186	0.248	0.122	0.150	0.130	0.159

Notes:*** p<0.01, ** p<0.05, * p<0.1 Robust standard errors in parentheses. All specifications control for province dummies. Parental education levels are controlled for in specifications 2,4,6,8 but estimated effects are not presented in this table.

Table 2.6: Estimated Effects of IVs on Years of Education, College Attendance and Log Annual Earnings

	(OLS Estimation) Dependent Variable: Years of Education			(Prboit Estimation) Dependent Variable:Collge Dummy			(OLS Estimation) Dependent (Variable: Log Annual Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Adm. Rate	0.017***		0.032***	0.014***		0.023***	0.003*		0.005***
	(0.004)		(0.005)	(0.004)		(0.005)	(0.002)		(0.002)
$NCEEHukou \times Rate$		-0.029***	-0.028***		-0.018***	-0.019***		-0.004**	-0.004***
		(0.004)	(0.004)		(0.004)	(0.004)		(0.002)	(0.002)
$NCEEHukou \times Tenure$		0.040*	0.041*		0.054***	0.051**		0.020**	0.019**
		(0.022)	(0.022)		(0.021)	(0.021)		(0.009)	(0.009)
$NCEEHukou \times Tenure^2$		-0.229***	-0.228***		-0.232***	-0.227***		-0.055**	-0.053**
		(0.066)	(0.066)		(0.061)	(0.060)		(0.026)	(0.026)
Polynomials(3) of NCEE-cohort	Yes		Yes	Yes		Yes	Yes		Yes
NCEE-cohort Dummies		Yes			Yes			Yes	
Exclusion Test(F/LR Test)	16.392	56.869	37.058	12.39	23.52	36.08			
Observations	3,724	3,724	3,724	3,724	3,724	3,724	3,724	3,724	3,724

Notes:*** p<0.01, ** p<0.05, * p<0.1 Robust standard errors in parentheses. Each specification includes male dummy, quadratical tenure, NCEE-Hukou dummy, number of siblings, urban dummy, province dummies and dummies indicating father's and mother's education levels. Estimates for their effects are not presented in this table. "Yes" indicates the controls are included.

Table 2.7: OLS and IV Estimates of Rate of Returns to Higher Education

	(OLS Estimation) Dependent Variable: Log Annual Earnings			(IV Estimation) Dependent Variable: Log Annual Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.072***	0.070***	0.071***	0.175*	0.150**	0.161***
	(0.007)	(0.007)	(0.007)	(0.097)	(0.060)	(0.051)
Male	0.210***	0.215***	0.211***	0.210***	0.215***	0.211***
	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.022)
Tenure	0.025***	0.016**	0.016**	0.020***	0.016**	0.015**
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$Tenure^2$	-0.030**	-0.015	-0.013	-0.027*	-0.022	-0.021
	(0.015)	(0.021)	(0.021)	(0.016)	(0.022)	(0.023)
$NCEEHukou \times Tenure$		0.020**	0.019**		0.014	0.012
		(0.008)	(0.008)		(0.010)	(0.010)
$NCEEHukou \times Tenure^2$		-0.040	-0.037		-0.021	-0.016
		(0.025)	(0.026)		(0.030)	(0.030)
NCEEhukou	-0.090***	-0.237***	-0.229***	-0.097***	-0.212***	-0.200***
	(0.027)	(0.056)	(0.056)	(0.028)	(0.061)	(0.060)
No. of Siblings	-0.031***	-0.030***	-0.030***	-0.022*	-0.023**	-0.022**
	(0.009)	(0.009)	(0.009)	(0.012)	(0.011)	(0.010)
Urban	0.177***	0.211***	0.206***	0.076	0.131*	0.114*
	(0.032)	(0.034)	(0.034)	(0.101)	(0.069)	(0.063)
Polynomials(3) of NCEE-cohort	Yes		Yes	Yes		Yes
NCEE-cohort Dummies		Yes			Yes	
Overidentification Test						0.026
						(0.872)
Exclusion Test(F-Statistics)				16.392	56.869	37.058
Observations	3,724	3,724	3,724	3,724	3,724	3,724
R-squared	0.221	0.229	0.224	0.166	0.197	0.182

Notes:*** $p<0.01$, ** $p<0.05$, * $p<0.1$ Robust standard errors in parentheses. Each specification includes province dummies and dummies indicating father's and mother's education levels. Estimates for them are not presented in this table. "Yes" indicates the controls are included.

Table 2.8: OLS and IV College Premium Estimates

	(OLS Estimation) Dependent Variable: Log Annual Earnings			(IV Estimation) Dependent Variable: Log Annual Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
College	0.306***	0.298***	0.301***	0.645***	0.506***	0.575***
	(0.031)	(0.031)	(0.031)	(0.137)	(0.144)	(0.138)
Male	0.211***	0.216***	0.212***	0.212***	0.217***	0.213***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Tenure	0.025***	0.017**	0.016**	0.021***	0.017**	0.016**
	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)
$Tenure^2$	-0.033**	-0.019	-0.017	-0.034**	-0.026	-0.026
	(0.015)	(0.021)	(0.021)	(0.015)	(0.021)	(0.021)
$NCEEHukou \times Tenure$		0.019**	0.018**		0.015*	0.013
		(0.008)	(0.008)		(0.009)	(0.009)
$NCEEHukou \times Tenure^2$		-0.038	-0.036		-0.026	-0.020
		(0.025)	(0.025)		(0.027)	(0.027)
NCEEhukou	-0.088***	-0.230***	-0.222***	-0.092***	-0.209***	-0.195***
	(0.027)	(0.056)	(0.056)	(0.028)	(0.058)	(0.058)
No. of Siblings	-0.033***	-0.033***	-0.032***	-0.029***	-0.030***	-0.029***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Urban	0.174***	0.208***	0.202***	0.095**	0.157***	0.134***
	(0.032)	(0.034)	(0.034)	(0.043)	(0.048)	(0.047)
Polynomials(3) of NCEE-cohort	Yes		Yes	Yes		Yes
NCEE-cohort Dummies		Yes			Yes	
Observations	3,724	3,724	3,724	3,724	3,724	3,724
R-squared	0.218	0.226	0.220	0.190	0.216	0.201

Notes:*** p<0.01, ** p<0.05, * p<0.1 Robust standard errors in parentheses. Each specification includes province dummies and dummies indicating father's and mother's education levels. Estimates for them are not presented in this table. "Yes" indicates the controls are included.

Chapter 3 Effects of Cohort Size on College Premium: Evidence from China's Higher Education Expansion

3.1 Introduction

As a leading proximate cause of rising overall earnings inequality since the 1980s in the U.S., the increase in the college/high school wage premium has been well documented. Authors such as Katz and Murphy (1992), Acemoglu (2002), and Autor et al. (2008) have explained the rise as the consequence of an accelerated rise in the relative demand for college graduates and an abrupt slowdown in the growth of the relative supply of college graduates.[1] These studies focus on the aggregate trend of the college wage premium that may conceal independent trends by age groups. Card and Lemieux (2001) argue that heterogeneous trends of college premium by age groups may arise if workers in different age groups within the same education group are imperfectly substitutable and the trends of the relative supply of college workers are heterogeneous by age groups. Using data from the United States, the United Kingdom and Canada, they demonstrate the imperfect substitution between age groups and attribute the observed relative rise in the college premium for younger workers since the early or mid 1980s to the stagnated growth of the relative supply of college educated workers among the young during the same periods.[2] The study in cohort size effect on relative earnings can be traced back to Freeman (1976). He expressed the concern that the decline in earnings of new-entrant college relative to high school graduates in the early 1970s signaled a long-run decline in the earning return to higher education due to the increasing number of college graduates. However, little evidence from other countries has been added until recently. Kawaguchi and Mori (2016) reveal the heterogeneous trends of the college premium by age groups between 1986 and 2008 in Japan. Our paper adds evidence to this literature by documenting the divergent trends of college premium by age groups between 1995 and 2013 in China, and examines how the college premium is affected by the age group specific relative size of college educated workers.[3]

In the two studies of the U.S., U.K., Canada, and Japan, an important identification issue arises, the relative size of the college educated population is likely responsive to the college premium. Identification typically rests upon exclusion restrictions for instruments. China presents a unique environment where the decision of

[1]It is argued that the increase may have been driven by both skill-biased technological change (SBTC) featured by the computer revolution and the outsourcing of manufacturing. Katz et al. (1999) and Autor et al. (2008) support the idea of SBTC, and Feenstra and Hanson (2001) support the idea of outsourcing. The growth of college graduation rates stagnated for cohorts born in the early 1950s and entered labor market in late 1970s. See Card and Lemieux (2001) for details.

[2]The relative rise in college premium for younger workers commenced 5 years later in the U.K. and Canada than in the U.S.

[3]Considering that there exists certain amount of workers below high school education in China, we focus on the college premium with respect to non-college workers. Results for the college/high

who obtains a college degree is determined by a national test. In most time periods, far more students take the test than are admitted. However, since 1977, the government expanded admissions and allowed additional students to enter college. Hence the Chinese experience embeds a natural experiment allowing for arguably exogenous determination of college attainment. Further, the identification for the four countries all rely on the relative rise in college premium for younger workers since early or mid 1980s and the associated relative slowdown in growth of relative supply of college workers among the young. This timing overlapped with the emergence of skill-biased technological change (SBTC) since the early 1980s with the onset of the computer revolution. And it is suggested that this computer driven technological change may increase the relative demand for college workers and further increase the college premium among the young in particular (Krueger, 1993; Card, 1999a; Freeman and Katz, 2007).[4] Therefore, the negative effect of age group specific relative size on age group specific college premium may have been confounded by SBTC and overestimated for the four countries. The distinct trends of college premiums and relative size of college workers during our study period of 1995 to 2013 in China allows for an probably underestimated magnitude of the negative cohort size effects. Finally, China is also worth examining due to its large population and workforce.

Using China Household Income Project (CHIP) 1995, 1999, 2002, 2007, and 2013, five repeated cross-sectional surveys, we find that the trends of the college premium between 1995 and 2013 by age groups are substantially different. In figure 3.1(a), the college premium as measured by log earnings ratio was very similar for younger (age 25-34) and older (age 45-54) groups, about 25 percentage points in 1995. As of 2013, the college premium for the younger group was about 30 percentage points, similar to the level in 1995, while the college premium for the older group was about 50 percentage points, nearly double that of 1995. In figure 3.1(b), we present the age group specific trends of the relative supply of college workers measured as log employment ratio. The relative supply for the younger group increased substantially while that for the older group was quite stable during the same period. Comparing these two figures, the stagnation of the college premium for the younger group between 1995 and 2013 was potentially due to the fast growing relative supply of college workers. Figures 3.2 and 3.3 show that in the U.S. and Japan, unlike in China, the college premium for the older group decreased with respect to the younger group while the relative supply for the older group increased with respect to the younger group.[5] If technological progress positively affects the college premium for the younger group particularly as the literature argues, the negative age group specific supply effects will be overestimated for the U.S. and Japan, and underestimated for China.

The underlying cause of the heterogeneous trends of relative supply by age groups

school premium will also be discussed and compared with existing studies.

[4]Card (1999a) uses relative computer usage rates of college workers as a proxy indicator of the relative complementarity of college workers with new technology and finds little evidence supporting this hypothesis. However, we have no evidence to reject the hypothesis and it may be argued that the proxy indicator may have failed to fully capture the relative complementarity.

[5]These two figures are taken from the paper by Kawaguchi and Mori (2016) who compare the

is the non-monotonic increase in the college attendance rate which was determined by college capacity and birth cohort size. The expansion of college attendance ended in 1965 in the U.S. and in 1975 in Japan.[6] Therefore, Card and Lemieux (2001) and Kawaguchi and Mori (2016) mainly study the post-expansion period for the U.S. and Japan.[7] In China, the growth in college attendance began in 1977 and did not slow down until 2008. This paper studies the period 1995-2013 which includes the expansion. Thus, this paper reveals the consequence of an ongoing college attendance expansion, supplementing previous studies on the consequence of past college attendance expansion.

In this paper, we follow the empirical strategy by Card and Lemieux (2001) to construct the college premium and relative supply by age and survey year, and to further regress the cell-specific college premium against the relative supply. The supply effect on the college premium is estimated to be about -0.1 by our main specification. That implies, when holding the age cohort and survey year constant, a one unit increase in the log relative size of college workers is associated with about 10 percentage points decrease in the college premium. The more comparable result, by focusing on the college/high school earnings premium, is about -0.18 which is slightly lower than -0.2 in the U.S. and -0.23 in the U.K. while almost same as the results for Japan and Canada. That the negative supply effect in China is so close to the other four countries is remarkable in view of the very different economic development levels, trends of the college premiums and the relative supply, and higher education expansion phases between China and the other four countries. It is more interesting considering that the estimate of the supply effect should be a lower bound in China and upper bound in the other four countries.

We further examine the heterogeneous supply effects by age groups and find that the entrant group between ages 25 and 29 is more substantially affected by their own relative supply. This finding can be used to address the ability composition issue.[8] The ability effect is argued to be more substantial for the older group (Lillard, 1977), however, the estimated negative supply effects for the older groups are significantly lower than that for entrant group. This implies that the ability composition effect is not a dominant part in the estimated supply effect even if it may exist to some extent.

The rest of this paper is organized as follows. Section 2 presents the theoretical model by Card and Lemieux (2001). Section 3 discusses empirical strategy and potential identification issues. Section 4 introduces our data from China and describes the trends of college/non-college earnings gap and relative supply of college workers with details. Section 5 presents main results and section 6 reports a set of robustness checks. Finally, we conclude in section 7.

---

trends between the U.S. and Japan.

[6] The fast growth in college attendance rate ended for U.S. birth cohort 1947 and Japanese birth cohort 1957 approximately (Kawaguchi and Mori, 2016). And suppose college age is 18.

[7] Even though the period studied by Card and Lemieux (2001) is from 1959 to 1996, the identification relies on data in years later than 1975.

[8] It is argued that the increase in relative supply of college workers might be associated with a

3.2 Theoretical Framework

Model Setup

We start with a Cobb-Douglas aggregate production function that has been widely used in the macro-growth literature:

$$Y_t = A_t L_t^{\alpha} K_t^{1-\alpha} \tag{3.1}$$

where subscript t indexes year, Y_t is aggregate output, A_t is total factor productivity, L_t is aggregate labor force input, K_t is physical capital input and α is the share of income allocated to labor force.

Following the existing literature on the trend of wage differentials by education (Katz and Murphy, 1992; Autor et al., 2008), we assume the labor force input L_t in equation 2.1 follows a CES aggregation of college and non-college laborïijŇ

$$L_t = [\sum_s (\theta_{st} L_{st}^{\rho})]^{1/\rho} \tag{3.2}$$

where subscript s indexes education level which takes c for college labor and n for non-college labor, θ_{st} is the technological efficiency parameter, and $-\infty < \rho \leq 1$ is a function of the elasticity of substitution σ_A between college and non-college labor force ($\rho = 1 - 1/\sigma_A$). The underlying assumption is that different age cohorts within the same education group are perfect substitutes. To explain the divergent trends of the college premiums across age cohorts, following Card and Lemieux (2001), we relax the assumption of perfect substitution across age cohorts and further assume the labor force of each education level is aggregated by age cohorts by CES functional formïijŇ

$$L_{st} = [\sum_j (\alpha_{sjt} L_{sjt}^{\eta_s})]^{1/\eta_s} \tag{3.3}$$

where subscript j indexes age cohort, α_{sjt} is a relative efficiency parameter,[9] $-\infty < \eta_s \leq 1$ is a function of the elasticity of substitution σ_s among different age cohorts ($\eta_s = 1 - 1/\sigma_s$), and L_{sjt} is size of labor force for each education-age-year group.

Profit-Maximizing Wage

In this setup, assuming efficient utilization of labor force, we can derive the profit-maximizing wage of an average worker with education level s, among age cohort j, in year t as the value of corresponding marginal productivity in log form:

$$log(w_{sjt}) = log(\Phi_t) + log(\theta_{st}) + (\frac{1}{\sigma_s} - \frac{1}{\sigma_A})log(L_{st}) + log(\alpha_{sjt}) - \frac{1}{\sigma_s}log(L_{sjt}) \tag{3.4}$$

---

decrease in the average ability gap that leads to a decrease in the college premium.(Chay and Lee, 2000; Taber, 2001; Juhn et al., 2005; Carneiro and Lee, 2009, 2011) Thus, the negative supply effect tends to be overestimated.

[9] This relative efficiency parameter may be affected by labor complementarity with technology,

where

$$\Phi_t = \alpha A_t K_t^{1-\alpha} L_t^{\alpha-\rho}$$

According to equation 2.4, the age specific variation in wages is due to the age specific variation in the relative efficiency parameter α_{sjt} and the size of labor force L_{sjt}. The term $log(\Phi_t)$ represents a common year fixed effect across education levels while the terms $log(\theta_{st}) + (\frac{1}{\sigma_s} - \frac{1}{\sigma_A})log(L_{st})$ represents the year fixed effect for specific education level s. In this setup, the coefficient of $log(L_{sjt})$, $-1/\sigma_s$, should be negative unless the labor forces are perfectly substitutable across age cohorts ($\sigma_s = \infty$).

Age Specific Relative Size and College Premium

It is straightforward to derive the college premium by taking difference of the log wages between college and non-college labor force in terms of equation 2.4,

$$log(\frac{w_{cjt}}{w_{njt}}) = log(\frac{\theta_{ct}}{\theta_{nt}})+(\frac{1}{\sigma_c}-\frac{1}{\sigma_A})log(L_{ct})-(\frac{1}{\sigma_n}-\frac{1}{\sigma_A})log(L_{nt})+log(\frac{\alpha_{cjt}}{\alpha_{njt}})-\frac{1}{\sigma_c}log(L_{cjt})+\frac{1}{\sigma_n}log(L_{njt}). \tag{3.5}$$

To simplify our explanation of the age specific college premiums, we assume that the extent of substitution across age cohorts is the same for the college and non-college labor force. That is, we assume $\eta_c = \eta_n = \eta$ (which is equivalent to $\sigma_c = \sigma_n = \sigma$). This assumption will be tested empirically. We can rewrite equation 2.5 as:

$$log(\frac{w_{cjt}}{w_{njt}}) = log(\frac{\theta_{ct}}{\theta_{nt}}) + (\frac{1}{\sigma} - \frac{1}{\sigma_A})log(\frac{L_{ct}}{L_{nt}}) + log(\frac{\alpha_{cjt}}{\alpha_{njt}}) - \frac{1}{\sigma}log(\frac{L_{cjt}}{L_{njt}}) \tag{3.6}$$

where $log(\frac{\theta_{ct}}{\theta_{nt}})$ implies the year trend of the relative technological efficiency for college labor force, $log(\frac{L_{ct}}{L_{nt}})$ measures the relative size of aggregate college labor fore in year t, $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$ is the age specific trend of relative efficiency of college workers, and $log(\frac{L_{cjt}}{L_{njt}})$ is the key variable of interest, the age specific relative size of college labor force.

Notice that the first two terms at the right-hand-side of equation 2.6 capture the year trend of the college premium common for all age cohorts. Thus, the heterogeneous trends of the college premium across age cohorts should be due to the last two terms. And, the negative effect of age specific relative size on the college premium is expected unless workers are perfectly substitutable across age cohorts (the substitution elasticity $\sigma = \infty$).

Birth Cohort Effects

The two age specific variables, $log(\frac{L_{cjt}}{L_{njt}})$ and $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$, are actually measures for the birth cohort $t-j$. Thus, in addition to a fixed age profile and year fixed effect, $log(\frac{L_{cjt}}{L_{njt}})$ should capture birth cohort effects that reflect the variation in college attendance rate

skill composition, ability composition, etc. Card and Lemieux (2001) assume the relative efficiency parameter is constant over time. In our paper, we relax the strict assumption to allow for time while $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$ should capture birth cohort effects that mainly reflect the technological changes. We can decompose them into age cohort, year, and birth cohort fixed effects,

$$log(\frac{L_{cjt}}{L_{njt}}) = F_{t-j} + F_j + F_t \tag{3.7}$$

$$log(\frac{\alpha_{cjt}}{\alpha_{njt}}) = f_{t-j} + f_j + f_t. \tag{3.8}$$

Therefore, we can rewrite equation 2.6 as

$$log(\frac{w_{cjt}}{w_{njt}}) = F'_t + F'_j + f_{t-j} - \frac{1}{\sigma}F_{t-j} \tag{3.9}$$

where

$$F'_t = log(\frac{\theta_{ct}}{\theta_{nt}}) + (\frac{1}{\sigma} - \frac{1}{\sigma_A})log(\frac{L_{ct}}{L_{nt}}) + f_t - \frac{1}{\sigma}F_t$$

$$F'_j = f_j - \frac{1}{\sigma}F_j.$$

This implies that the college premium for age cohort j in year t can be decomposed into year, age and birth cohort fixed effects. Only in the case that workers are not perfectly substitutable across age cohorts ($\sigma < \infty$) can birth cohort effects in relative size, F_{t-j}, contribute to the birth cohort fixed effects in the college premium.

3.3 Empirical Approach

Construction of College Premium and Relative Cohort Size

Our primary goal in this paper is to estimate the effect of the age cohort specific relative size of college workers on the age cohort specific college premium. Since these two key variables are not directly observed in our data set, we need to construct measures of them prior to further analysis.

Following the standard approach in the literature on cohort size effects, we collapse individual data into cells based on single-year age and survey year. Then the age specific college premium in each survey year is estimated with the individual observations within corresponding cells by following specification,

$$y_i = \beta_0 + \beta_1 college_i + \varepsilon_i \tag{3.10}$$

where y_i is log annual earnings, β_0 is a constant, $college_i$ is a dummy variable that takes 1 for college workers and 0 for non-college workers, and β_1 is the college premium to be estimated. Some existing papers (Welch, 1979; Card and Lemieux, 2001; Brunello, 2010) on the effect of cohort size on earnings or the college premium use log weekly or hourly wages for analysis. However, in terms of equations 2.4 and 2.5, we believe that using weekly or hourly earnings is inappropriate unless the age specific

variation which will be helpful to explain potential identification issues.

relative size is measured using total working weeks per year or total working hours per year correspondingly. Due to the lack of information of working hours, we use log annual earnings for our analysis.

Accordingly, we build the measure of age specific relative size based on the number of workers.[10] The age-year cell specific relative size is just the log ratio of the number of college workers to the number of non-college workers within each cell.

Following Card and Lemieux (2001), we also record the standard errors of estimated cell specific college premiums. The corresponding inverse variances will be used as weights for the regression analysis to put more weight on those precisely estimated college premiums, and be used to construct goodness-of-fit tests for the null hypothesis that the relevant specification has no specification error.[11]

To improve the precision of the estimated college premiums and to reduce the sampling variation in relative size of college workers, we construct cells based on three-year age and survey year alternatively at the expense of reducing the number of cells for regression analysis by two thirds. Nevertheless, this serves as a good robustness check.

Testing the Assumption: Equally Substitutable College and Non-College Labor

In section 2.3, we link age specific college premiums to age specific relative sizes by equation 2.6 based on the assumption that the substitution elasticity across age cohorts, σ_s, is the same among college and non-college groups. It is a hypothesis that needs to be tested. Following the profit-maximizing wage equation 2.4 for an average worker in age cohort j with education level s in year t, we decompose the unobserved three-way variable $log(\alpha_{sjt})$ into three two-way fixed effects (education level-year, education-age, and age-year fixed effects) and a conditional zero mean error term ε_{sjt}. Then we test the assumption by OLS estimation with the following specification:

$$log(w_{sjt}) = F_{st} + F_{sj} + F_{jt} + \beta_1 noncollege_s \times log(L_{sjt}) + \beta_2 college_s \times log(L_{sjt}) + \varepsilon_{sjt} \tag{3.11}$$

where the dependent variable $log(w_{sjt})$ is mean log earnings for j years old workers with education level s in year t, the education-year fixed effects F_{st} absorbs the terms $log(\Phi_t) + log(\theta_{st}) + (\frac{1}{\sigma_s} - \frac{1}{\sigma_A})log(L_{st})$ from equation 2.4 and the additional education-year fixed effect decomposed from $log(\alpha_{sjt})$, the education-age fixed effect F_{sj} captures the potentially different age-profile of earnings for college and non-college groups, the

[10]Using annual earnings and number of workers to build measures for the college premium and relative size highlights that our estimated effects of cohort size on the college premium have slight different implications from those using weekly earnings or hourly earnings. Considering that working hours or working weeks are endogenously determined in the labor market, using them to measure relative size may suffer the identification issue of reverse causation.

[11]Essentially, it tests whether the recorded variances of the estimated college premiums are significantly different from the variances of the residual in relevant specification. See Card and Lemieux

age-year fixed effect F_{jt} captures those unobserved factors that commonly affect both education groups, and $log(L_{sjt})$ is the age cohort size for education group s in year t. We allow for a different effect of cohort size on earnings by including the interaction terms between education group dummy and age cohort size, $college_s \times log(L_{sjt})$ and $noncollege_s \times log(L_{sjt})$. We test whether $\beta_1 = \beta_2$.

An equivalent test strategy as follows is based on equation 2.5,

$$log(\frac{w_{cjt}}{w_{njt}}) = F_t + F_j + \beta_1 log(L_{cjt}) + \beta_2 log(L_{njt}) + \varepsilon_{jt} \tag{3.12}$$

where dependent variable is estimated college premium for age cohort j in year t, the age-year fixed effect in equation 3.2 is canceled out by taking difference between log earnings of college workers and non-college workers. Noticing that β_1 and β_2 represent $-\frac{1}{\sigma_c}$ and $\frac{1}{\sigma_n}$ respectively, we test if $\beta_1 + \beta_2 = 0$.

Since both dependent variables in equations 3.2 and 3.3 are estimated first, the corresponding standard error can been obtained prior to the tests. Following the literature, we use inverse squared standard errors as weights to implement weighted-OLS estimation.

Estimating the Effect of Age Specific Relative Size on College Premium

Our basic specification to estimate the effect of age specific relative size on the college premium is based on the equation 2.6. We decompose the unobserved age-year log ratio of relative efficiency, $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$, into age fixed effect, year fixed effect and age-year two-way variation. We use the following specification,

$$r_{jt} = F_t + F_j + \beta_1 log(\frac{L_{cjt}}{L_{njt}}) + \varepsilon_{jt} \tag{3.13}$$

where r_{jt} is the estimated college premium for age cohort j in year t, F_t captures all year specific factors, F_j is the age fixed effect decomposed from $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$, $log(\frac{L_{cjt}}{L_{njt}})$ is relative size of college workers measured as log ratio of the number of college workers to the number of non-college workers within each age-year cell, and the error term ε_{jt} is assumed to be conditional zero mean to ensure the OLS estimate of β_1 identifies the relative size effect on the college premium.

However, a simple OLS estimate of β_1 may be biased through two ways. First, our specification is strictly based on the profit-maximizing wage functions which reflect only the demand side of the labor market, whereas the observed college premiums and age specific relative sizes represent the realized general equilibrium. Therefore, $log(\frac{L_{cjt}}{L_{njt}})$ may have been affected by the college premium through a supply channel. We use the predetermined variable, age-year cell specific log ratio of the number of college degree holders to the number of non-college degree holders (including both employed and unemployed individuals) as an instrumental variable for $log(\frac{L_{cjt}}{L_{njt}})$.

(2001) for details.

Second, the error term ε_{jt} captures not only those plausible zero mean sampling error and specification error, but also the age-year two-way variation from the unobserved log relative efficiency ratio, $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$. The simple OLS estimate of β_1 will be biased due to omission of relevant variables if $log(\frac{L_{cjt}}{L_{njt}})$ is correlated with the unobserved two-way varying $log(\frac{\alpha_{cjt}}{\alpha_{njt}})$. By the implication of the relative efficiency parameter α, we know it may be affected by relative labor complementarity with technology, relative skill composition, relative ability composition, etc. Since it has been discussed that the skill biased technological change favoring younger college workers allows for a lower bound of the estimates in the context of China, we focus on the potential ability composition effect and skill composition effect in this section.

Ability Composition Effect

It's widely believed that basic OLS estimates of college premium are biased due to unobserved ability or self-selection, which is reflected by the huge literature on isolating the returns to college from the returns to ability. However in the literature on the evolution of college premium, the change in the ability composition effect receives much less attention. Some studies find that the changes in ability composition or self-selection indeed contribute to the observed college premium evolution, even if the extents are found to be different (Chay and Lee, 2000; Taber, 2001; Juhn et al., 2005; Carneiro and Lee, 2009, 2011).[12]

Before presenting our empirical strategy to address the ability composition effect, it is necessary to explain how it may confound the estimate of the relative size effect in this paper. As we noted in section 2.4 and which will be empirically explored, the relative size $log(\frac{L_{cjt}}{L_{njt}})$ captures strong birth year fixed effects which drive the age-year two-way variation in $log(\frac{L_{cjt}}{L_{njt}})$. There has been an observed increase in college attainment along with the birth cohorts. And the observed increase stems from both demographic changes, and an expanding capacity of China's higher education. In China's strict test score-based college admission system, it's plausible that marginal college students have lower ability than the average college students. When the expansion of college capacity outpaced the demographic changes in China, the share of college students increased, marginal students entered college, and the average ability of college students was lowered. By the same logic, the average ability of non-college students also has been lowered. The lowered average abilities for both education groups result in difficulty in predicting the sign of the correlation between relative size $log(\frac{L_{cjt}}{L_{njt}})$ and relative average ability. However, some previous papers show that the ability effect on earnings for high school graduates is insignificant (Carneiro and Lee, 2011) and is less positive than that on college graduates (Lillard, 1977; Carneiro and Lee, 2011). This evidence implies that we should be careful that the negative correlation between relative size of college workers and the earnings gap effect of relative average ability may lead our estimated relative size effect on earnings gap to

[12] Among these studies, only Carneiro and Lee (2011) focus on isolating the ability composition

be downward biased. In the extreme case, what we estimated for β_1 by equation 3.4 may just be an ability composition effect rather than a relative size effect.

Our strategy is to explore the age pattern of the potentially confounded relative size effect by allowing for heterogeneity across age groups,

$$r_{jt} = r_t + r_j + \beta Agp_j \times log(\frac{L_{cjt}}{L_{njt}}) + \varepsilon_{jt} \tag{3.14}$$

where Agp_j is a vector of age group dummies, β is the corresponding vector of coefficients which captures the relative size effects on college premium across age groups, and ε_{jt} is suspected to include ability composition effects negatively correlated with $log(\frac{L_{cjt}}{L_{njt}})$. If the ability composition effects are significant and indeed negatively correlated with $log(\frac{L_{cjt}}{L_{njt}})$, by simple OLS estimation, we will obtain an estimated age group pattern of relative size effect dominated by the age group pattern of ability composition effects.

Lillard (1977) uses NBER-Th data[13] which includes measured ability (AFQT scores) and reveals that the earnings effect of measured ability increases as one ages and this increasing pattern is more significant for college graduates than for high school graduates.[14] More specifically, the ability effect is almost negligible or even slightly negative under age 35 and peaks around age 50. Taking this pattern as also true in China's context,[15] the estimated relative size effects will be more negative for older groups if the ability composition effects exist and are negatively correlated with $log(\frac{L_{cjt}}{L_{njt}})$. Therefore, if an opposite pattern is revealed by our estimation, we will be able to argue that the confounded ability composition effects are trivial, and the estimated effects for younger groups, especially those under age 35, should be uncontaminated by the ability composition effects at least. The opposite pattern can be explained as the younger groups tend to be affected by their own cohort relative size more substantially.[16]

Skill Composition Effects

We use occupation and industry composition to capture the skill composition approximately. The variation in age specific relative size, $log(\frac{L_{cjt}}{L_{njt}})$, is mainly driven by China's higher education expansion since 1977 when the national college entrance ex-

---

effect within the age specific framework as we do in this paper, while others within aggregate framework.

[13]NBER-Th sample was based on a sample of men who had volunteered for pilot, bombardier, and navigator programs of the Air Force during World War II. Thomas Juster organized a resurvey of a subset of these men in 1969 and built a data set providing information on education, income, AFTQ test scores and detailed information on various measures of family background.

[14]One explanation is that the more able tend to invest more in on-job training or choose more promising jobs.

[15]Even if there is no evidence from China's data, we believe the underlying logic also holds in China's labor market.

[16]Welch (1979) finds that the cohort size effects are more negative for entrant cohorts with data

amination was restored. One year later, in 1978, China started "the open and reform" through which China switched from a central-planned economy to market-oriented economy gradually. Along with the transition, new labor market entrants with different education levels may have been reallocated into occupations and industries differently. Considering that the higher education expansion and economy transition took place during the same period, it is possible that the age-year variations in occupation and industry compositional differential between college and non-college groups are correlated with the age-year variation in college/non-college relative size. That means, in equation 3.6, the omitted occupation and industry compositional effects are possibly correlated with $log(\frac{L_{cjt}}{L_{njt}})$. Due to sample size limitation,[17] we are not able to control for these compositional effects consistently for each age-year cell. Therefore, we turn to regression with individual data directly by the following specification,

$$\begin{aligned} y_{ijt} = \beta_0 + \beta_1 college_{ijt} \times log(\frac{L_{cjt}}{L_{njt}}) + \beta_2 log(\frac{L_{cjt}}{L_{njt}}) \\ + F_t + F_j + college_{ijt} \times (F_t + F_j) + \gamma X_{ijt} \times F_t + \varepsilon_{ijt} \end{aligned} \tag{3.15}$$

where i, j, t denotes individual, and X_{ijt} includes a series of dummies for occupation and industry categories. We allow for the occupation and industry fixed effects vary across years by the interaction term $X_{ijt} \times F_t$. With this specification, the OLS estimate of β_1 is the relative size effect on the college premium conditional on occupation and industry. Dropping the interaction term $X_{ijt} \times F_t$ should result in an estimated β_1 close to those by specification 3.4 since the earnings gap by specification 3.6 can be expressed in the exactly same form:

$$E[Y_{ijt}|college_{ijt} = 1] - E[Y_{ijt}|college_{ijt} = 0] = \beta_1 log(\frac{L_{cjt}}{L_{njt}}) + F_t + F_j. \tag{3.16}$$

By controlling for these labor market destinations, we also alleviate another concern about the college majors composition effect since it is plausible that majors determine college graduates' occupation and industry destinations to a substantial extent.[18]

3.4 Data

Our data are drawn from five repeated cross-section nationally representative surveys - China Household Income Project (CHIP) 1995, 1999, 2002, 2007 and 2013.[19] As

of the U.S.

[17]On average, in our data set, each age-year cell contains about 90-210 individuals.

[18]Grogger and Eide (1995) reveal that the trend away from low-skill subjects such as education and toward high-skill subjects such as engineering accounts for one-fourth of the rise in the male college wage premium with the U.S. data. Majors information is not available in our data set that we can't directly control for them.

[19]CHIP 2008 surveys the same individuals in 2007, so we pool them together and notate it as CHIP2007 in this paper.

indicated by its name, CHIP surveys detailed household income, education, employment, and family background information, which makes it a widely used data source in the literature on earnings differential across education or other labor market-related topics in China.[20] In this paper, following the literature (Zhang et al., 2005; Ge and Yang, 2011; Wang, 2012; Wang et al., 2014) on China's college premium, we focus on the urban samples.[21] We further restrict our sample to males between 25-54. Only focusing on males avoids the selection issue due to intermittent female labor force participation.[22] The lower limit, age 25, is to make sure most college graduates have entered the labor market while the upper limit, age 54, is to drop those near retirement age who may decide to retire non-randomly (Brunello, 2010).

We define individuals who have a three-year college degree, a four-year college degree or above as college graduates, and all other individuals as non-college graduates. This broad definition has the advantage of covering all workers in the labor market and obtaining more precise estimates for earnings gaps by keeping more observations, but the disadvantage of bringing the contamination of composition effects. Therefore, we will also present results based on only 4-year college and high-school graduates as a robustness check.

We use annual earnings to estimate the college premiums due to limited consistent information on working weeks and hours. However, CHIP (2007) only provides monthly earnings information without working months available. Fortunately, the potential inconsistence in estimated college premiums for wave 2007 should be captured by a fixed year effect which will be controlled for in our empirical analysis.

We collapse the individuals between age 25-54 into 150 cells based on single-year age and survey year. For each cell, our estimated college premiums and the key explanation variable, relative size of college workers are further based on those employed individuals reporting positive annual earnings. The instrumental variable for relative size of college workers, as discussed in section 3.3, is based on both employed and unemployed individuals between 25-54, including females. This wide inclusion is to make sure we construct a predetermined variable only affected by the exogenous demographic change and higher education expansion.

Sample Summary

Before presenting a graphical description of cell-specific relative size and estimated college premium, we summary our filtered sample in table 3.1. The number of observations in each survey year ranges between 2754 and 6461 and the variation is mainly due to the variation in sample size of original surveys. The average log annual

[20]For instance, Gustafsson et al. (2008) write a whole book using CHIP to explore inequality and public policy in China.

[21]The main reasons documented are that rural household income is generally indivisible, there is a relatively low share working in non-agriculture sectors, and there are few college graduates working in rural area.

[22]See Card and Lemieux (2001) and Brunello (2010). Even if this issue may not be as severe as that in western countries considering that female labor force participation is relatively high in

earnings shows steady increase.[23] The share of college workers increased from 29% in 1995 to 45% in 2007 and drops slightly to 42% in 2013, even if the higher education expansion should have pushed up the college share. This reflects that men's share of college workers in urban areas has achieved a saturation level and more young college graduates have to stay in rural areas.[24] The age structure is stable during the covered period shown by the stable averages and standard deviations. By categorizing occupations into three levels (high-skill, mid-skill, and low-skill levels), we can see a decrease in high-skill share and increase in low-skill share.[25] Most industry shares are stable, except that manufacturing share decreased while service shares increased. The dominant industry by share of employment changed from manufacturing to service. As we discussed in section 3.3.2, if these changes in occupation and industry shares were different between education groups and age groups, our estimated effect of the relative size on the college premium would be contaminated by occupation and industry compositional effects.

Relative Sizes and Estimated College Premiums

For each age-year cell, we can estimate a college premium by equation 3.1, and measure the corresponding relative size of workers as the log ratio of the number of college workers to the number of non-college workers. Figure 3.4 provides pairs of these two variables. Due to the year fixed effects and the intrinsic age profile, it shows no clear linear relationship between the college premium and the relative size of college workers. Nevertheless, figure 3.4 reveals substantial variations in the two variables, which makes it possible for us to identify the potential relationship by regression analysis.

By exploring the changing age profiles of college premium and relative size, we can reveal the relationship between them graphically. To make sure our graphs suffer less estimation variation, we use 30 broader cells of five-year age and survey year. Figures 3.5 and 3.6 present the age profiles of the college premium and the relative size respectively across survey years. As the downward age profile of the relative size turned much steeper form 1995 to 2013 in figure 3.6, the age profile of the college premium departed from flat pattern to a upward pattern in figure 3.5. The opposite switching age profiles of relative size and the college premium is a reflection of the negative relationship.

China(Meng, 2012), we focus on males for comparing results with existing literature mainly on western countries.

[23]We use nominal annual earnings in this paper, so the increase captures both real income growth and inflation. Using nominal earnings does not affect our results since the inflation index is canceled out in the estimates of college/non-college earnings gap.

[24]By comparing the share of college graduates in rural area between 2007 and 2013 using CHIP rural surveys, we indeed find this trend.

[25]High-skill level includes principals and professional technicians, mid-skill level includes clerical/office staff and low-skill level includes the other occupations.

Relative Size, College Premium and Higher Education Expansion

As we discussed in section 2.4, the relative size for college workers in age cohort j and year t is measuring those born in year t-j, which implies that it should have captured strong birth year effects in addition to a fixed age profile and year fixed effect. To graphically illustrate the birth cohort effects, we plot the share of college workers against birth year groups in figure 3.8. Even if the profiles shifts up and down across years and may also have absorbed intrinsic age structure, it is clearly revealed that there are steady rises in the share of college workers from birth year group 1953-1958 to 1984-1988. Considering that high school students usually take the national college entrance examination (NCEE) at about 18 years old, the rising birth year trends coincide with the restored NCEE and the expansion of higher education since 1977 as figure 3.7 shows.[26] The positive correlation implies that the rise in relative size of college workers across birth years was mainly driven by the higher education expansion.

We also check if the college premiums also show strong birth year fixed effects, which would serve as preliminary evidence of the effect of the relative size on the college premium as we discussed in section 2.4. Due to the more substantial variations in the college premium across years and age cohorts, the graph for the college premium suffers greater noise than the graph for shares of college workers. Therefore, we turn to regressions based on equations 2.7 and 2.9 which decompose relative size and college premium for age cohort j in year t into year, age and birth cohort fixed effects.

Table 3.2 presents results of the decompositions. We take survey year 1995 and birth group 1941-1958 as reference groups.[27] In column 1, we decompose college premiums by basic OLS estimation. In column 2 we weight our regression by the inverse sampling variance of estimated college premium with the χ^2 statistic for testing specification error reported.[28] Since the results are just different slightly between basic OLS and Weighted OLS estimation, we focus on the weighted-OLS results following the literature. Year fixed effects on the college premium increased by 38.3 percentage points from 1995 to 2013, and about half of the increase happens between 1995 and 1999. The estimated birth year fixed effects show a steady decreasing trend for those born after 1958. Specifically, comparing with those born in 1941-1958, the college premium for the recent birth cohorts 1984-88 decreased by almost 39 percent. As the χ^2 static 111.07 and its p-value 0.45 indicate, we fail to reject the null hypothesis that there is no specification error in our model. The dependent variable in column 3 is the share of college workers while the dependent variable in column 4 is relative size of college workers which is also the explanatory variable in our main specification 3.4

[26]This figure depicts the nationwide trend including both urban and rural while figure 3.8 is based on CHIP's urban samples only. The absolute shares of college workers are much higher than those in figure 3.8. This implies that more college students are from urban areas or stay in urban areas.

[27]Considering that most high school students apply for college at about 18 years old, those born before 1958 arrived at college age before 1977 when the NCEE was restored. We do not divide our sample evenly into birth groups due to the uneven year gaps of our surveys.

[28]The null hypothesis is that there is no specification error conditional on included fixed effects.

to be estimated in next section. Estimated year fixed effects capture both sampling variation and overall relative employment across survey years. As the results in column 3 show, comparing with 1995 conditionally, about 3.5 percent more college workers were employed in 1999 and 9.6 percent less college workers were employed in 2013. The estimated birth year fixed effects show a steady rising trend for those born after 1958, which reveals a negative correlation with the estimated fixed effects on college premium in column 2. The predicted birth group fixed effects on the share of college workers and college premium, standardized to age 40 and year 2013, are plotted in figures 3.9 and 3.10. The contrasting trends together with the higher education expansion in figure 3.7 provide preliminary evidence that higher education expansion drove the rise in share of college workers which further compressed the college premium.

By exploring the decomposed birth year fixed effects on the two key variables, we can find that their age-year two-way variations are mainly captured by the birth cohort fixed effects and our identification of the effect of relative size on earnings gap relies just on these two-way variations. Therefore, if any other birth cohort specific factors affecting college premium are correlated with the birth cohort specific variation in relative size of college workers, our identification of the relative size effect will fail. As we discussed in section 3.4, the main contaminating factors are potentially correlated compositional effects due to the birth cohort specific variations in ability, occupation and industry compositions.

3.5 Results

In table 3.3, we present our basic estimates of the effect of age specific relative size of college workers on the college premium based on specification 3.4 which regresses the age specific college premium against age and year fixed effects and the age specific relative size of college workers. The results by weighted/unweighted OLS estimation in columns 1 and 2 do not show significant differences. The estimated effects of the relative size of college workers on the college premium, -0.08 and -0.078 are quite similar and significant at the 5% level. They imply that, holding year and age constant, a one unit increase in the relative size of college workers leads to about 8 percentage points decrease in the college premium. By the model implication, these estimates represent that the elasticity of substitution across age cohorts is about 12.5. The estimated year fixed effects show that the college premium increased steadily until 2007 and then fell slightly in 2013, which indicates that the macro conditions may have favored college workers relatively during the covered period.

As we discussed in section 3.4, basic OLS estimation may suffer the issue of simultaneous causation which makes it biased. We use the predetermined variable, log ratio of the number of college graduates to the number of non-college individuals (including both male and female, employed and unemployed), as an instrumental variable for our independent variable based only on male workers. The corresponding

---

See Card and Lemieux (2001) for details.

results are presented in column 3 and 4. The magnitudes of the estimated relative size effects increase by about 30 percent, even if these increases are not statistically significant. The slightly attenuated OLS estimates imply that the relative size of college workers might be positively affected by the college premium simultaneously. In other words, higher college premium induces relatively more college graduates to seek employment, which is consistent with basic intuition even if this is not empirically studied in this paper.

However, our results above may still suffer bias due to omission of relevant variables as we discussed in section 3.4, such as ability, occupation and industry compositional factors which may be correlated with the relative size of college workers. To address the potential ability compositional effects, we explore the age group pattern of the relative size effect on the college premium based on equation 3.5. The corresponding results are presented in table 3.4. In column 1 of table 3.4, we divide ages into 6 groups evenly: 25-29, 30-34, 35-39, 40-44, 45-49 and 50-54. The estimated effects are significant only for the new entrants between age 25 and 29, -.142, at the 1% level. Thus, we alternatively divide ages into two groups, new entrants 25-29 and all other ages 30-54. Corresponding results are presented in column 2. The estimated effect for the new entrant group is still negative and significant, -0.156, while for all other ages is insignificantly negative, at -0.049. The F statistic implies that the effects are different significantly at the 5% level. The magnitudes of IV estimates in column 3 increase slightly, which reveals a similar pattern that new entrants are more substantially affected by their own relative size than the older group (age 30-54). If our estimates are dominated by the ability composition effect, the revealed pattern should be the opposite showing a smaller negative effect for new entrants because the conditional ability effects are more substantial for older workers by Lillard (1977) as we discussed in section 3.4.1. Our estimated pattern is also consistent with the findings by Welch (1979) that entrant cohorts are more easily affected by the cohort size effect. As Welch (1979) argues, in the early career phase, workers as learners accumulate skills gradually. Due to the substantial variance of the skills possessed, entrant workers are less easily substituted with each other, therefore, more easily affected by their own cohort size. As they enter later career phases and accumulate enough skills to fulfill different tasks, they are more substitutable and less easily affected by the cohort size.

In the specifications for our main findings above, we define the college premium and relative size of college workers based on broadly defined college workers including three-year college graduates or above and corresponding non-college graduates. We believe this definition has the advantage in covering all workers in the labor market and keeping as many observations as possible to obtain precisely estimated college premium for further analyzing the relative size effect on the college premium. However, the estimated college premium by our definition is different from the college premium referring to the earnings gap between 4-year college and high-school graduates, which leads our analysis to be less relevant to the huge literature on college premiums and less comparable to several studies on the effect of relative size on college premium (Card and Lemieux, 2001; Carneiro and Lee, 2009; Kawaguchi and Mori, 2016). Another disadvantage is that the potential varying average years of schooling for broadly defined college and non-college groups may bring in additional sources of variation in the estimated college premium.[29]

Therefore, we measure relative size of college workers and estimate college premium based on the sample including only four-year college workers and high-school workers. Results are presented in table 3.5. The magnitudes of our OLS and IV estimates presented in columns 1 to 4 increase slightly but the increases are not significant compared with the results by the broader definition of college and non-college. To make our results more comparable with Card and Lemieux (2001) using data from the U.S., U.K., and Canada, we follow their method for measuring relative size. They use the college premium (earnings gap between 4-year college and high-school) as the dependent variable while using a relative size measure based on all education levels as independent variable.[30] We follow their measure for relative size notated as LRS in table 3.5. The estimated effect, -0.178, in column 5 is much larger by magnitude than -0.101 in column 1 and becomes very similar to the results by Card and Lemieux (2001), -0.203 for the U.S., -0.233 for U.K. and -0.165 for Canada.[31] That the negative supply effect in China is so close to these three countries is remarkable in view of the very different economic development levels, trends of the college premiums and the relative supply, and higher education expansion phases between China and the other three countries. It is more interesting considering that the estimate of the supply effect should be a lower bound in China and upper bound in the other three countries.

3.6 Robustness Checks

In this section, we first test the underlying assumption of our main specification 3.4. After presenting the positive results for the assumption that college workers and non-college workers are equally substitutable across age cohort, we use several alternative specifications to check the robustness of the effect of age specific relative size of college workers on the age specific college premium.

Testing the Assumption: Equal Education-Specific Elasticity of Substitution

As we discussed in section 2.3, to directly link the relative size of college workers and the college premium like equation 2.6 entails the assumption of identical elasticity of substitution across age cohorts for college and non-college groups. The testing results are presented in table 3.6. In column 1, we estimate a model based on equation

[29] The average year of schooling for non-college group increased substantially because of family income growth and China's nine-year compulsory education program implemented since 1985.

[30] To account for differences in the effective labor supply by different education levels, they also assign a weight to each level with the average earnings. However, we have to point that they use hourly wage rates and annual working hours to construct their college premium and relative size. In our data, information about working hours is not available.

[31] The larger estimated absolute effects by this alternative measure LRS comes from its high

3.2 without controlling for the age-year two-way fixed effects. The estimated effect of college workers' size on college workers' average earnings is significantly negative, -0.146, while that for non-college workers is insignificantly positive, 0.04. By the high F statistic with nearly zero p-value, we have to reject the null hypothesis of identical effects. However, we can reject the hypothesis of no specification error at the 1% level as the corresponding χ^2 statistic indicates. After we control for the age-year two-way fixed effects in the specification for column 2, we find that the age-specific size effects for college workers and non-college workers are similar, and we can't reject the null hypothesis of identical effects by the corresponding F statistic, 0.41 with p-value 0.522. Meanwhile, the χ^2 statistic testing the hypothesis that there is no specification error reduces substantially from 409.86 in column 1 to 120.64 with p-value 0.313. The comparison implies that there exists a common age-year fixed effect on average earnings for both college and non-college workers. In column 3, the equivalent specification to that for column 2 is based on equation 3.3, which leads to estimates with almost identical magnitudes. The opposite signs of the estimated effects are consistent with the model implication since the dependent variable is the estimated college premium instead of education-specific average earnings. The corresponding F and χ^2 statistics have large p-values, which indicates that we can't reject the null hypothesis of identical effects and the null hypothesis of no specification error.

Controlling for Occupation and Industry

To deal with the potential confounding factors due to occupation and industry compositions, we directly control for these factors with individual data based on equation 3.6. Results are presented in table 3.7. In columns 1 and 2, we present results without controlling for occupation or industry as a comparison with the results by structural specifications in which these composition effects are not controlled for. As expected, we obtain very similar results of the effects of relative size on the college premium. The estimated average effect over all ages is -0.074 in column 1, while the effect is -0.191 for entrant group and -0.044 for older group in column 2. After controlling for year-varying fixed effects of occupation, industry and province, the results change slightly and the changes are not significant. This implies that these suspected confounding composition effects are not a serious issues. In columns 5 and 6 we present IV estimates which are very similar with corresponding estimates in tables 3.3 and 3.4.

Results for Women only and Pooled Women and Men

Focusing on men only is only appropriate conditional on a strong assumption that men and women in the same age cohort, education level, and survey year are not substitutable. Therefore, we first replicate our analysis for women only and then for pooled women and men under the assumption that men and women in same age

correlation with the basic measure and its smaller variation. A one unit change in this alternative measure is associated with about 1.5 units change in the basic measure.

cohort, education level and survey are perfect substitutable. For the sake of brevity, we only present OLS estimates in table 3.8. The results with women only in columns 1 and 2 are not only smaller by magnitude but also less precise than those with men only while the results with both men and women are very similar. Another interesting finding comes from the difference in year trends between women and men. Comparing the estimated fixed year effects in column 1 of table 3.8 and column 2 in table 3.3, we can find that men's college premium increased more rapidly than women's from 1995 to 2013.

Several Other Specification Checks

We have performed several other specification checks of which the results are presented in table 3.9.

Firstly, we notice that CHIP 1999 and 2007 draw samples from provinces that are partially different from those in CHIP 1995, 2002 and 2013 even though each wave is nationally representative. Therefore, it is naturally to check the robustness using CHIP 1995, 2002 and 2013 only to keep the province composition constant.[32] The corresponding results are presented in columns 1 and 2.

Secondly, by checking individual's rural-urban migration status, we find that the proportion of rural-urban migrants increased steadily from about 18 percent in 1999 to about 32 percent in 2013.[33] Considering that including rural-urban migrants may introduce an added source of variation in the college premium due to endogenous self-section, we focus on those non-migrants to check the robustness of relative size effect on the college premium and present the results in columns 3 and 4.

Lastly, to reduce sampling variations, we also construct broader cells based on three-year age and survey year at the expense of reducing number of cells by two thirds, from 150 to 50. The corresponding results are presented in columns 5 and 6.

Even if the OLS and IV estimates in columns 1-4 are less precise than our previous main results due to the drop of CHIP 1995 (or both 1995 and 2007), their magnitudes are similar. The results with broader cells shown in column 5 and 6 are significantly negative with similar magnitudes. Overall, these alternative specifications show robust results of the relative size effects on college premium.

3.7 Conclusion

In this paper, we document the divergent trends of the college premiums across age groups from 1995 to 2013 in China. Comparing with the well-studied increasing overall trend during the same period, this divergence has received little attention. Specifically, the college premium in 2013 for the younger group (age 25-34) was about

[32]Even though we have controlled for province fixed effects in our previous specification with individual data, we perform the estimation with structural model as a double check.

[33]We define those born with rural residence registration changed to urban residence registration. In CHIP 1995, we can't accurately identify the migration status that we only use the waves 1999, 2002, 2007 and 2013

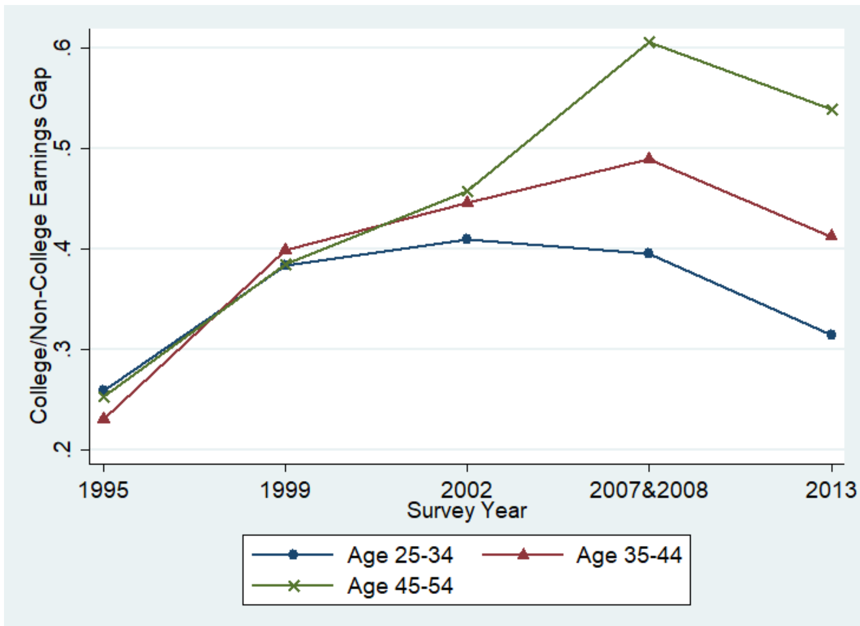
30 percentage points, similar to the level in 1995, while the college premium in 2013 for the older group (age 45-54) increased to 50 percentage points nearly double that of 1995. To attribute these divergent trends of college premium to the changes in relative size of college workers, we use the model by Card and Lemieux (2001) which incorporates imperfect substitution between similarly educated workers in different age cohorts. Due to the distinctions of these trends in China, our identification is free of the overestimation issue due to the technological progress which possibly favored younger college workers in particular. Our results are similar to those in the U.S., U.K., Canada, and Japan. Holding the age cohort and survey year constant, a one unit increase in relative size of college workers is associated with about 10 percentage points decrease in college/non-college premium and about 18 percentage points decrease in college/high school premium. That the negative supply effect in China is so close to the other four countries is remarkable in view of the very different economic development levels, trends of the college premiums and the relative supply, and higher education expansion phases between China and the other four countries.

We further find that the negative effect is much more substantial among the new entrants (age 25-29) than among the experienced workers (age 30-54). By this pattern, we not only demonstrate that the new labor market entrants are more sensitive to their own cohort relative size but also argue that the confounding ability composition effect should not be a serious issue.

Figures

Figure 3.1: Trends of College Premium and Relative Supply of College Workers by Age Groups: China

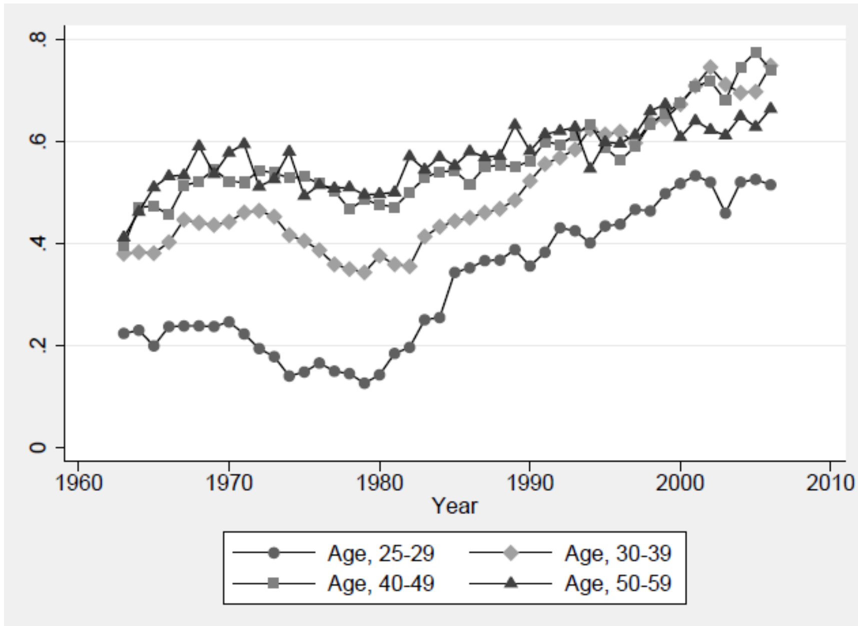
(a) Log College/Non-College Earnings by Age Groups



(b) Log College/Non-College Supply by Age Groups

Figure 3.2: Trends of College Premium and Relative Supply of College Workers by Age Groups: The U.S.

(a) Log College/HS Wage by Age Groups



(b) Log College/HS Supply by Age Groups

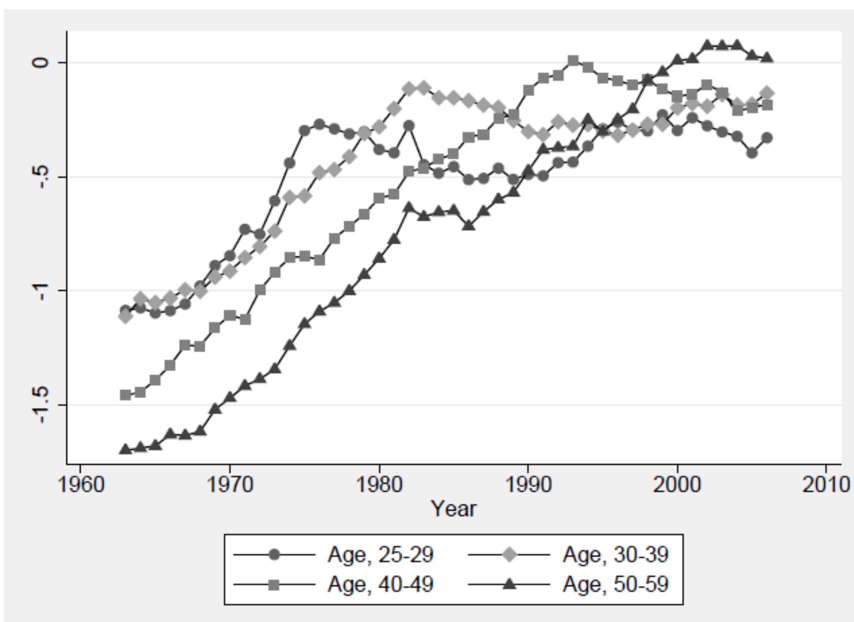
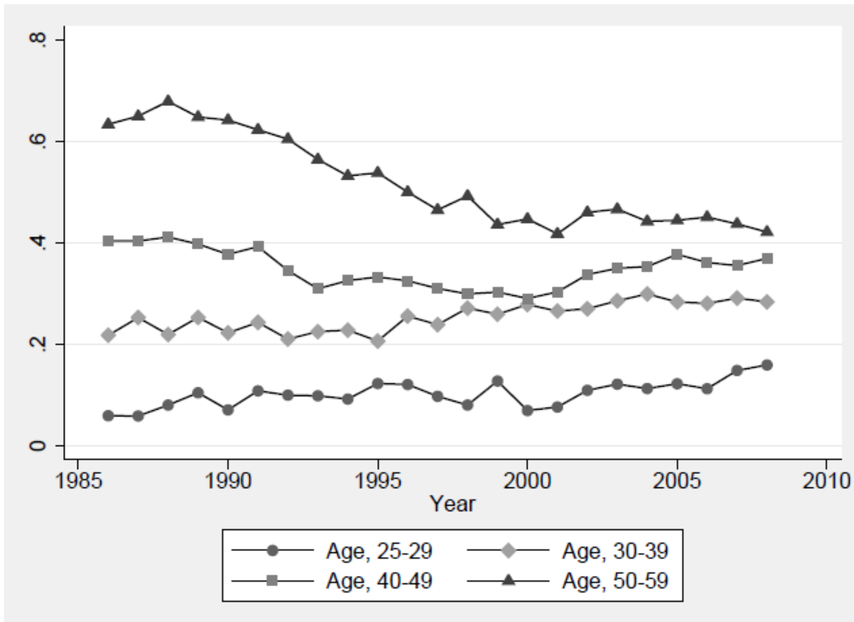


Figure 3.3: Trends of College Premium and Relative Supply of College Workers by Age Groups: Japan

(a) Log College/HS Wage by Age Groups



(b) Log College/HS Supply by Age Groups

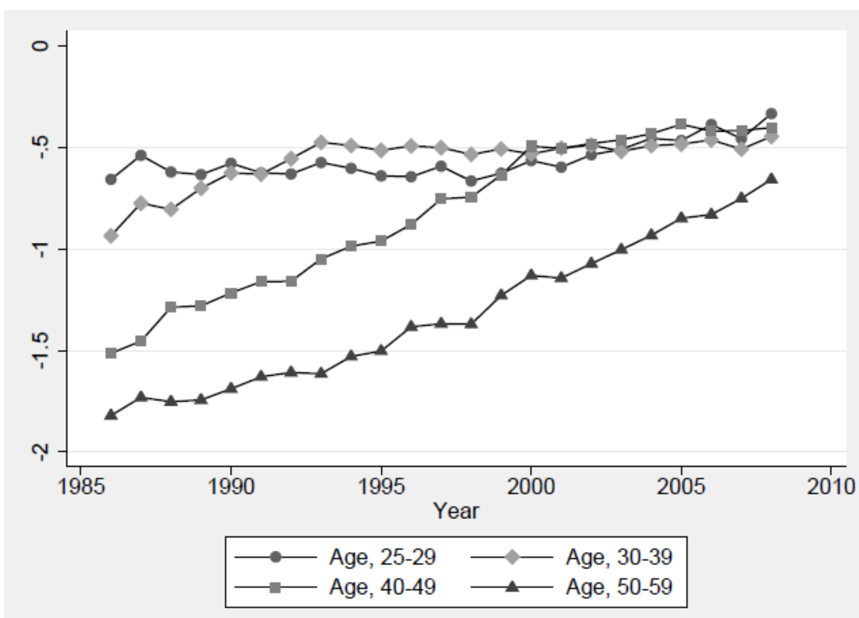


Figure 3.4: Age-Year Cell Specific Log Relative Sizes and Estimated College Premiums

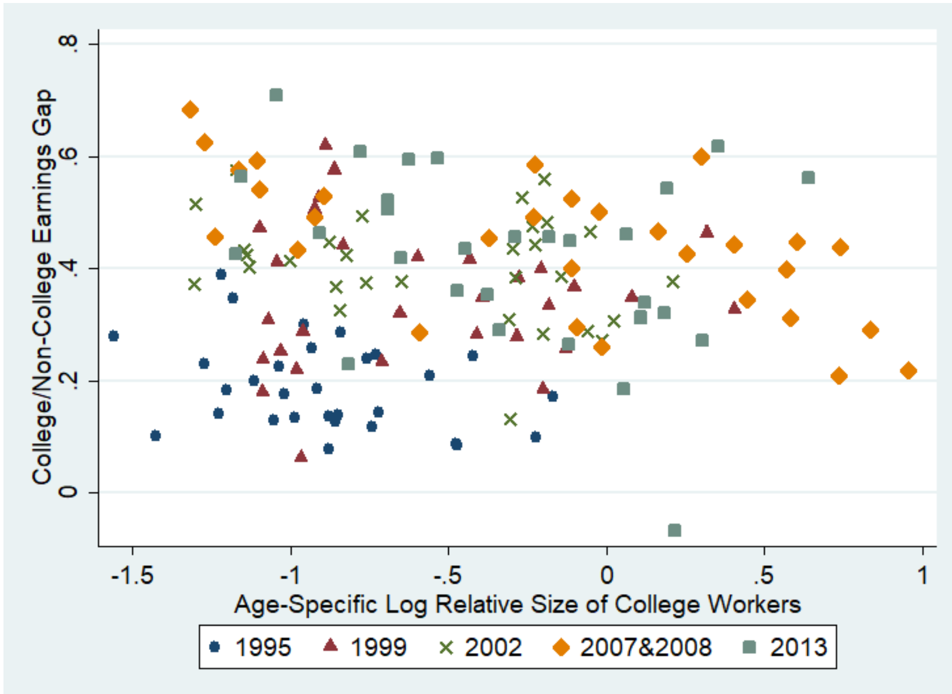


Figure 3.5: Male Workers' Age Profiles of the College Premium Across Years

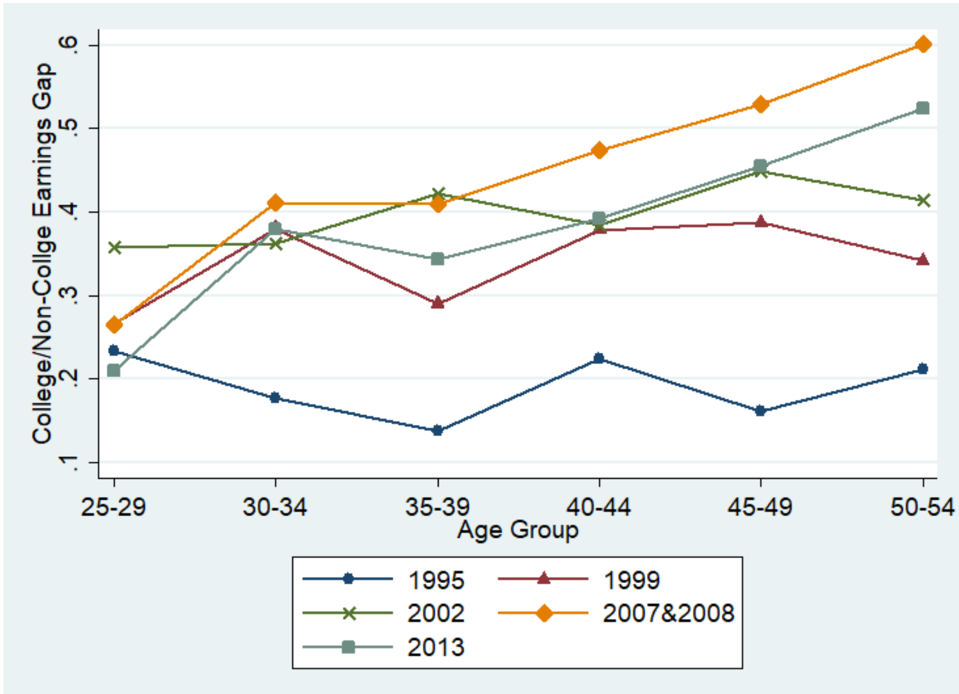


Figure 3.6: Male Workers' Age Profiles of Relative Size Across Years

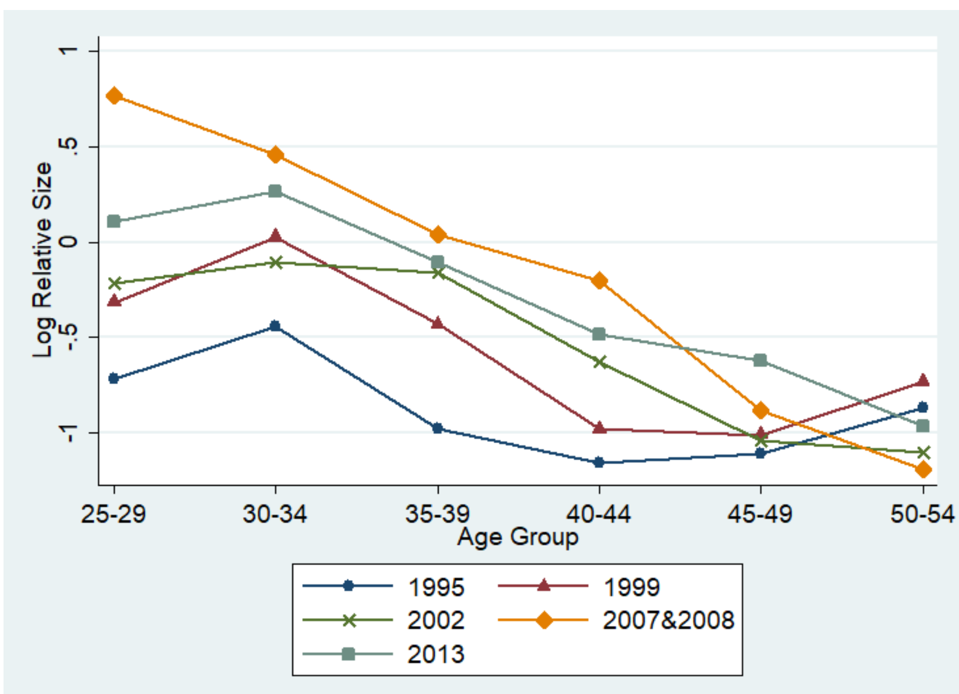


Figure 3.7: Demographical Change and Higher Education Expansion in China

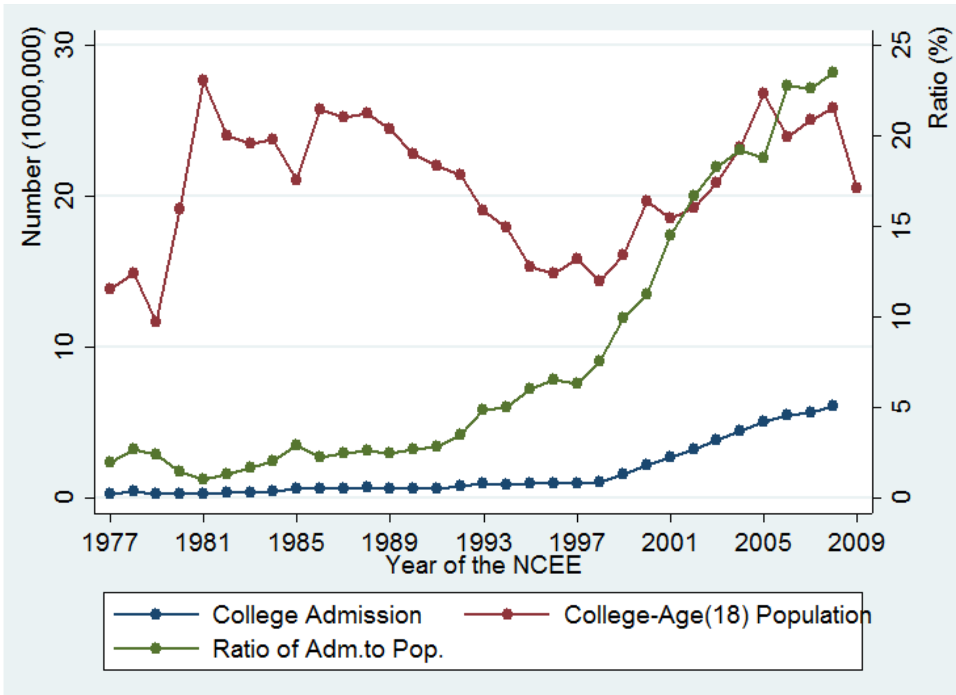


Figure 3.8: Birth Year Profiles of the Share of College Workers

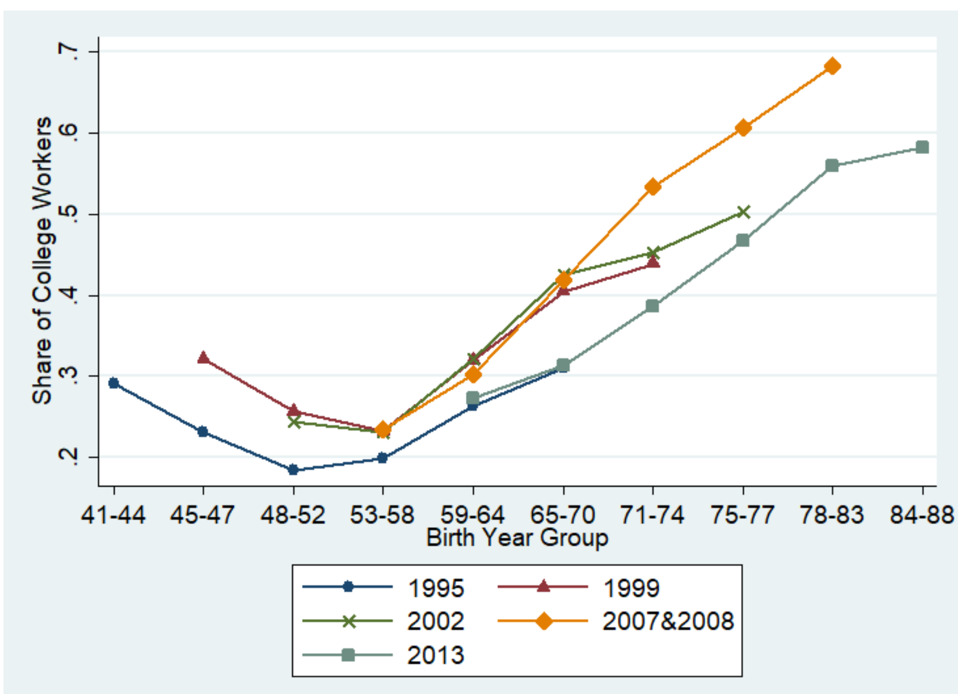


Figure 3.9: Predicted Birth Group Fixed Effects on the College Premium

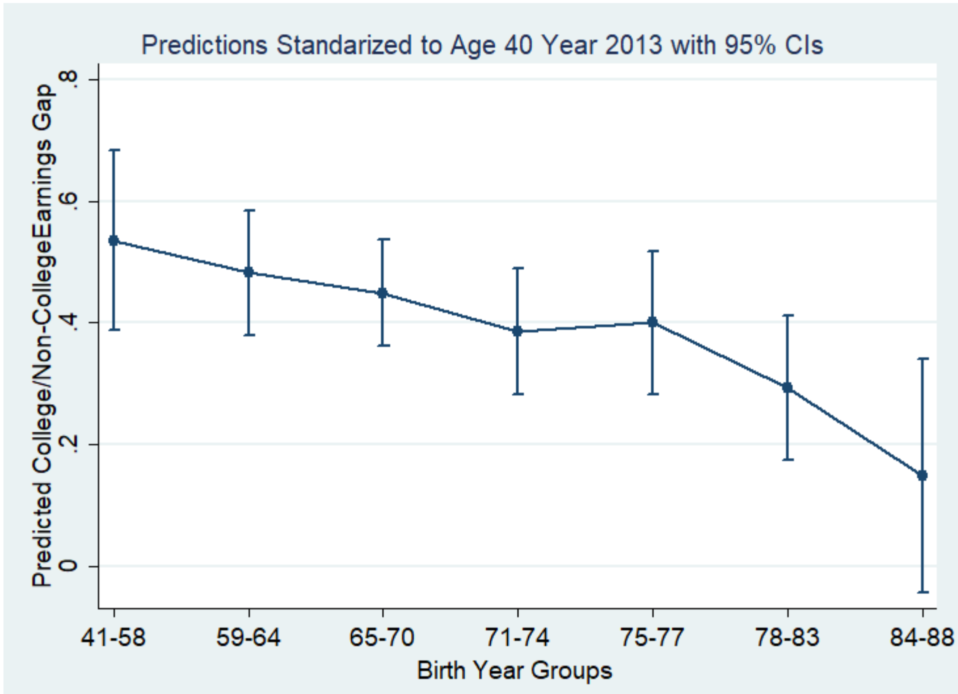
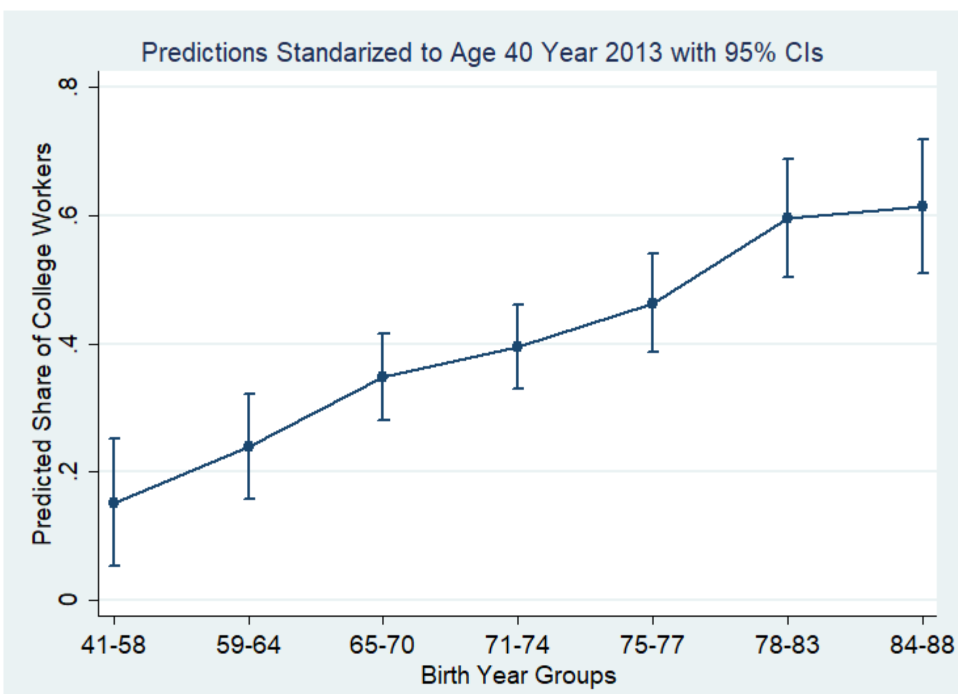


Figure 3.10: Predicted Birth Group Fixed Effects on the Share of College Workers



Tables

Table 3.1: Summary Statistics: Male Workers Only

CHIP	1995	1999	2002	2007	2013
Average Annual Earnings (Yuan)	6,374	8,690	11,159	27,174	36,316
College	0.29	0.35	0.36	0.45	0.42
	(0.45)	(0.48)	(0.48)	(0.50)	(0.49)
Age	39.84	40.59	41.41	40.48	40.74
	(7.70)	(7.52)	(7.62)	(8.28)	(8.18)
High-Skill Occ.	0.40	0.41	0.40	0.34	0.25
	(0.49)	(0.49)	(0.49)	(0.47)	(0.43)
Mid-Skill Occ.	0.20	0.16	0.18	0.22	0.18
	(0.40)	(0.37)	(0.38)	(0.41)	(0.39)
Low-Skill Occ.	0.40	0.43	0.42	0.45	0.57
	(0.49)	(0.50)	(0.49)	(0.50)	(0.50)
Agriculture	0.02	0.01	0.01	0.01	0.02
	(0.14)	(0.11)	(0.11)	(0.10)	(0.14)
Mining	0.01	0.04	0.03	0.01	0.04
	(0.11)	(0.18)	(0.17)	(0.10)	(0.21)
Construction	0.03	0.05	0.04	0.05	0.07
	(0.18)	(0.22)	(0.20)	(0.21)	(0.26)
Manufacturing	0.43	0.32	0.27	0.20	0.15
	(0.49)	(0.47)	(0.45)	(0.40)	(0.36)
Transportation etc.	0.06	0.16	0.14	0.16	0.14
	(0.24)	(0.37)	(0.35)	(0.37)	(0.35)
Trade	0.12	0.08	0.10	0.11	0.10
	(0.33)	(0.27)	(0.30)	(0.31)	(0.30)
Finance	0.05	0.03	0.04	0.08	0.06
	(0.22)	(0.17)	(0.19)	(0.28)	(0.24)
Service	0.13	0.20	0.23	0.29	0.28
	(0.34)	(0.40)	(0.42)	(0.45)	(0.45)
Public Institutions	0.14	0.11	0.14	0.09	0.13
	(0.35)	(0.32)	(0.34)	(0.29)	(0.33)
Observations	4978	2754	4900	6461	4335

Table 3.2: Birth Year Fixed Effects on College Premium and Relative Size

	(1) College Premium	(2) College Premium	(3) College Share	(4) Log Relative Size
Year Fixed Effects:				
1999	0.187***	0.183***	0.035**	0.159**
	(0.031)	(0.032)	(0.016)	(0.072)
2002	0.257***	0.252***	0.014	0.059
	(0.031)	(0.028)	(0.016)	(0.073)
2007	0.351***	0.349***	0.020	0.064
	(0.045)	(0.043)	(0.026)	(0.119)
2013	0.399***	0.383***	-0.096**	-0.427**
	(0.067)	(0.062)	(0.038)	(0.175)
Birth Fixed Effects:				
1959-64	-0.066	-0.053	0.088***	0.416***
	(0.043)	(0.038)	(0.025)	(0.114)
1965-70	-0.112*	-0.086	0.196***	0.894***
	(0.062)	(0.056)	(0.033)	(0.151)
1971-74	-0.184**	-0.149**	0.244***	1.098***
	(0.079)	(0.073)	(0.043)	(0.199)
1975-77	-0.151*	-0.135	0.312***	1.389***
	(0.090)	(0.085)	(0.051)	(0.236)
1978-83	-0.283***	-0.243**	0.443***	1.950***
	(0.108)	(0.098)	(0.064)	(0.292)
1984-88	-0.441***	-0.388***	0.462***	2.033***
	(0.150)	(0.137)	(0.076)	(0.349)
χ^2(p-value)		111.07(0.45)		
Observations	150	150	150	150
R-squared	0.943	0.951	0.985	0.910

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference year is 1995, reference birth group is 1941-1958. Age fixed effects are not shown. Weights used in specification 2 are inverse variances of estimated college premiums.

Table 3.3: Basic Estimates for Effects of Age Specific Relative Size of College Workers on College Premiums

Dependent Variable: College Premium	(1) OLS	(2) Weighted-OLS	(3) IV	(4) Weighted-IV
Log Relative Size	-0.080**	-0.078***	-0.111***	-0.103***
	(0.032)	(0.030)	(0.032)	(0.029)
Year Effects:				
1999	0.187***	0.188***	0.197***	0.195***
	(0.029)	(0.031)	(0.026)	(0.027)
2002	0.245***	0.246***	0.256***	0.254***
	(0.027)	(0.023)	(0.024)	(0.020)
2007	0.316***	0.328***	0.338***	0.345***
	(0.031)	(0.030)	(0.029)	(0.028)
2013	0.277***	0.293***	0.295***	0.307***
	(0.032)	(0.031)	(0.028)	(0.028)
F Statistic			447.17	655.04
χ^2(p-value)		115.85(0.46)		113.28(0.53)
Observations	150	150	150	150
R-squared	0.938	0.949	0.938	0.949

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The dependent variable for all specifications is the college premiums by age and year. All specifications also include age fixed effects not reported. The instrumental variable for log relative size is log ratio of the number of college degree holders (including both male and female, employed and unemployed) to the number of non-college degree holders. Weights for specifications in columns 2 and 4 are the inverse sampling variance of estimated college premiums. Reference year is 1995.

Table 3.4: Heterogeneous Relative Size Effects across Age Groups

Dependent Variable: College Premium	(1) Weighted-OLS	(2) Weighted-OLS	(3) Weighted-IV
Log Relative Size:			
Age 25-29 (New Entrants)	-0.142***		
	(0.050)		
Age 30-34	0.007		
	(0.060)		
Age 35-39	0.015		
	(0.055)		
Age 40-44	-0.064		
	(0.059)		
Age 45-49	-0.075		
	(0.090)		
Age 50-54	-0.172		
	(0.099)		
Age 25-29 (New Entrants)		-0.156***	-0.190***
		(0.047)	(0.044)
Age 30-54		-0.049	-0.069**
		(0.033)	(0.033)
F statistic(p-value)		4.51(0.04)	6.62(0.01)
χ^2(p-value)	107.74(0.54)	111.73(0.54)	109.41(0.52)
Observations	150	150	150
R-squared	0.953	0.951	0.951

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variables for all specifications are estimated college premiums by age and year. All specifications also include age and year fixed effects not reported. The instrumental variable for log relative size is log ratio of the number of college degree holders (including both male and female, employed and unemployed) to the number of non-college degree holders. All specifications are weighted by the inverse sampling variance of estimated college premiums.

Table 3.5: The Results to Sample including only High-School and 4-Year College Workers

Dependent Variable: College Premiums	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) OLS	(6) OLS
Log Relative Size	-0.101***		-0.128***			
	(0.028)		(0.028)			
Log Relative Size (Age 25-29)		-0.219***		-0.225***		
		(0.045)		(0.040)		
Log Relative Size (Age 30-54)		-0.081***		-0.099***		
		(0.029)		(0.029)		
LRS (Alternative Measure)					-0.178***	
					(0.043)	
LRS (Age 25-29)						-0.323***
						(0.067)
LRS (Age 30-54)						-0.147***
						(0.045)
Year Fixed Effects:						
1999	0.257***	0.252***	0.264***	0.257***	0.267***	0.258***
	(0.040)	(0.039)	(0.036)	(0.035)	(0.039)	(0.040)
2002	0.319***	0.318***	0.325***	0.321***	0.358***	0.351***
	(0.033)	(0.032)	(0.029)	(0.028)	(0.035)	(0.035)
2007	0.560***	0.565***	0.589***	0.582***	0.588***	0.580***
	(0.048)	(0.048)	(0.045)	(0.045)	(0.050)	(0.052)
2013	0.359***	0.353***	0.383***	0.369***	0.399***	0.386***
	(0.043)	(0.044)	(0.039)	(0.040)	(0.048)	(0.050)
Observations	150	150	150	150	150	150
R-squared	0.929	0.933	0.928	0.933	0.930	0.932

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variables for all specifications are the estimated college premiums by age and year. All specifications also include age fixed effects not reported. In column 3 and 4, the instrumental variable for log relative size is log ratio of the number of all college degree holders (including both male and female, employed and unemployed) to the number of non-college degree holders. Weights for specifications in columns 2 and 4 are the inverse sampling variance of estimated college premiums. To compare with the result by Card and Lemieux (2001), the alternative measure for log relative size, LRS in columns 5 and 6 is constructed based on all education levels rather than only high-school and 4-year college. Reference year is 1995.

Table 3.6: Testing Assumption: Identical Elasticity of Substitution for College and Non-College Workers

	(1) Average Earnings	(2) Average Earnings	(3) College Premium
Log Size of	-0.146***	-0.096**	0.093***
College Workers	(0.029)	(0.037)	(0.035)
Log Size of	0.040	-0.067	-0.062
Non-College Workers	(0.031)	(0.042)	(0.039)
F Statistic testing	28.95	0.41	0.47
"Identical Effects"	(0.000)	(0.522)	(0.495)
χ^2 Statistic testing	409.86	120.64	115.39
"No Specification Errors"	(0.000)	(0.313)	(0.446)
$Age \times Year$ Fixed Effects	NO	YES	NO
Observations	300	300	150
R-squared	0.989	0.997	0.949

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions are weighted by the inverse sampling variance of the corresponding dependent variable. Specifications in column 1 and 2 also include a set of year and age effects fully interacted with college dummy variable. Specification in column 3 also includes age and year fixed effects.

Table 3.7: Results Using Individual Data Controlling for Province, Occupation and Industry

Dependent Variable: Log Annual Earnings	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV
$College \times LogRelativeSize$	-0.074**		-0.083***		-0.106***	
	(0.031)		(0.028)		(0.031)	
$College \times LogRelativeSize$ (Age 25-29)		-0.191***		-0.177***		-0.208***
		(0.049)		(0.044)		(0.047)
$College \times LogRelativeSize$ (Age 30-54)		-0.044		-0.057*		-0.069**
		(0.034)		(0.031)		(0.034)
$LogRelativeSize$	0.166***	0.296***	0.144***	0.243***	0.184***	0.281***
	(0.019)	(0.034)	(0.017)	(0.031)	(0.019)	(0.034)
$LogRelativeSize \times$ 1[Age 30-54]		-0.170***		-0.128***		-0.128***
		(0.036)		(0.033)		(0.036)
(Province,Occupation,Industry)$\times$Year	NO	NO	YES	YES	YES	YES
Observations	23,428	23,428	23,428	23,428	23,428	23,428
R-squared	0.573	0.573	0.653	0.653	0.653	0.654

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All specifications also include college, age, and year fixed effects, the interaction between college and year fixed effects, and the interaction between college and age fixed effects.

Table 3.8: Robustness of The Results to Female Sample and Pooled Sample Including both Male and Female

Dependent Variable: College Premiums	(1) Women Only	(2) Women Only	(3) Men and Women	(4) Men and Women
Log Relative Size	-0.044		-0.071**	
	(0.042)		(0.028)	
Log Relative Size (Age 25-29)		-0.128***		-0.161***
		(0.045)		(0.033)
Log Relative Size (Age 30-54)		-0.016		-0.034
		(0.041)		(0.026)
Year Fixed Effects:				
1999	0.132***	0.124***	0.175***	0.165***
	(0.039)	(0.037)	(0.023)	(0.022)
2002	0.178***	0.172***	0.226***	0.216***
	(0.042)	(0.041)	(0.023)	(0.023)
2007	0.243***	0.233***	0.307***	0.295***
	(0.055)	(0.053)	(0.030)	(0.029)
2013	0.205***	0.193***	0.273***	0.255***
	(0.056)	(0.054)	(0.030)	(0.030)
Observations	150	150	150	150
R-squared	0.955	0.957	0.972	0.975

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variables for all specifications are estimated college premiums by age and year. All specifications also include age fixed effects not reported. All specification are weighted by the inverse sampling variance of the estimated college premiums.

Table 3.9: Robustness of The Results to Several Alternative Specifications

Dependent Variable: College Premium	(1) Same Provinces	(2) Same Provinces	(3) Non-Migrants Only	(4) Non-Migrants Only	(5) 3-Age-Year Cells	(6) 3-Age-Year Cells
Panel A: OLS Estimates:						
Log Relative Size	-0.069		-0.069		-0.101***	
	(0.045)		(0.043)		(0.026)	
Log Relative Size (Age 25-29)		-0.290**		-0.132*		-0.158***
		(0.117)		(0.071)		(0.030)
Log Relative Size (Age 30-54)		-0.045		-0.041		-0.077**
		(0.041)		(0.052)		(0.031)
R-squared	0.955	0.959	0.956	0.956	0.986	0.987
Panel B: IV Estimates:						
Log Relative Size	-0.090**		-0.113**		-0.109***	
	(0.046)		(0.045)		(0.022)	
Log Relative Size (Age 25-29)		-0.302***		-0.191***		-0.172***
		(0.089)		(0.061)		(0.029)
Log Relative Size (Age 30-54)		-0.053		-0.074		-0.084***
		(0.042)		(0.055)		(0.026)
R-squared	0.955	0.959	0.955	0.956	0.986	0.987
Observations	90	90	120	120	50	50

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The dependent variables for all specifications are estimated college premiums for by age and year (or by 3-age and year). All specifications also include age and year fixed effects not reported. All specification are weighted by the inverse sampling variance of the estimated college premiums. The instrumental variable for log relative size is log ratio of the number of college degree holders (including both male and female, employed and unemployed) to the number of non-college degree holders.

Chapter 4 Education-Occupation Match and the Trends in the College Earnings Premium by Age Groups in Urban China, 1995-2013

4.1 Introduction

As a leading proximate cause of rising overall earnings inequality since the 1980s in the U.S., the increase in the college/high school wage premium has been well documented. Authors such as Katz and Murphy (1992), Acemoglu (2002), and Autor et al. (2008) have explained the rise as the consequence of an accelerated rise in the relative demand for college graduates and an abrupt slowdown in the growth of the relative supply of college graduates.[1] These studies focus on the aggregate trend of the college wage premium that may conceal independent trends by age groups. Card and Lemieux (2001) argue that heterogeneous trends of college premium by age groups may arise if workers in different age groups within the same education group are imperfectly substitutable and the trends of the relative supply of college workers are heterogeneous by age groups. Using data from the United States, the United Kingdom and Canada, they demonstrate the imperfect substitution between age groups and attribute the observed relative rise in the college premium for younger workers since the early or mid 1980s to the stagnated growth of the relative supply of college educated workers among the young during the same periods.[2] Kawaguchi and Mori (2016) reveal the heterogeneous trends of the college premium by age groups between 1986 and 2008 in Japan. The second essay of my dissertation adds evidence to this literature by documenting the divergent trends of college premium by age groups between 1995 and 2013 in China. All these studies attribute the divergent trends in college premiums across age groups to the negative cohort size effect. As the relative supply of college educated workers increases, in a competitive labor market, college workers' average relative earnings which equal the relative marginal product decrease. However, no studies have examined the channel through which the cohort relative supply affect the college premium.

China's higher education expansion has led to an increasing number of college-educated workers, especially among the younger cohorts. More and more young workers can be placed into occupations which have typically been held by people with less education and lower earnings. It's possible that the cohort relative supply of college educated workers negatively affects the college premium by rematching workers with different education levels to different occupations. This paper examines how much the education-occupation match accounts for college earnings premium and

[1] It is argued that the increase may have been driven by both skill-biased technological change (SBTC) featured by the computer revolution and the outsourcing of manufacturing. Katz et al. (1999) and Autor et al. (2008) support the idea of SBTC, and Feenstra and Hanson (2001) support the idea of outsourcing. The growth of college graduation rates stagnated for cohorts born in the early 1950s and entered labor market in late 1970s. See Card and Lemieux (2001) for details.

[2] The relative rise in college premium for younger workers commenced 5 years later in the U.K.

to what extent the trends in education-occupation match contribute to the divergent trends in college earnings premium across age groups in urban China between 1995 and 2013. Using Blinder-Oaxaca decomposition (Oaxaca, 1973; Blinder, 1973), it's found that, for all survey years and age groups, the differences in the share of higher-skilled occupations between college and non-college educated workers only explains a small part of college premium, 10%-30%. The part due to the occupational premium is negligible. Over 70% of the college premium is contributed by the college premium for the workers with lower-skilled occupations. These unexpected results reveal that the relative supply of college educated workers among certain age groups has a general effect rather than age group specific effect on the occupation reallocation.

Since the college premium among lower-skilled occupations account for over 70% the overall college premium and is the driven force of the divergent trends in the college premium by age groups, to explain this result, this paper further develops the model by Card and Lemieux (2001) to include one more labor force aggregation step at the occupation level. Using the specification based on the derived result, the cohort size effect on college premiums among the lower-skilled and higher-skilled occupations can be estimated separately. Holding the age group and survey year constant, a one unit increase in log relative size of college educated workers is associated with about 15.3 percentage points decrease in college earnings premium among lower-skilled occupations, and 9.9 percentage points decrease among higher-skilled occupations.

The rest of this paper is organized as follows. Section 2 summarizes the data. Section 3 discusses decomposition method and results. Section 4 develops the model by Card and Lemieux (2001) to estimate the effect of cohort size on college premium among specific occupation category. Finally, I conclude in section 5.

4.2 Data Description

The data are drawn from five repeated cross-section nationally representative surveys - China Household Income Project (CHIP) 1995, 1999, 2002, 2007 and 2013.[3] CHIP surveys detailed household income, education, employment, and family background information, which makes it a widely used data source in the literature on earnings differential across education or other labor market-related topics in China.[4] In this paper, following the literature (Zhang et al., 2005; Ge and Yang, 2011; Wang, 2012; Wang et al., 2014) on China's college premium, I focus on the urban samples.[5] I further restrict the sample to males between 25-54. Only focusing on males avoids

---

and Canada than in the U.S.

[3]CHIP 2008 surveys the same individuals in 2007, so I pool them together and notate it as CHIP2007 in this paper.

[4]For instance, Gustafsson et al. (2008) write a whole book using CHIP to explore inequality and public policy in China.

[5]The main reasons documented are that rural household income is generally indivisible, there is a relatively low share working in non-agriculture sectors, and there are few college graduates working in rural area.

the selection issue due to intermittent female labor force participation.[6] The lower limit, age 25, is to make sure most college graduates have entered the labor market while the upper limit, age 54, is to drop those near retirement age who may decide to retire non-randomly (Brunello, 2010).

The individuals who have a three-year college degree, a four-year college degree or above are defined as college graduates, while all other individuals as non-college graduates. This broad definition has the advantage of covering all workers in the labor market and obtaining more precise estimates for earnings gaps by keeping more observations, but the disadvantage of bringing the contamination of composition effects. Therefore, I will also present results based on only 4-year college and high-school graduates as a robustness check.

The annual earnings are used to estimate the college premiums due to limited consistent information on working weeks and hours in the data. However, CHIP (2007) only provides monthly earnings information without working months available. Fortunately, the potential inconsistence in estimated college premiums for wave 2007 should be captured by a fixed year effect which will be controlled for in the empirical analysis. Samples are further restricted to those employed individuals reporting positive annual earnings.

Table 4.2 summarizes the data. The number of observations in each survey year ranges between 2754 and 6461 and the variation is mainly due to the variation in sample size of original surveys. The average log annual earnings shows steady increase.[7] The share of college workers increased from 29% in 1995 to 45% in 2007 and drops slightly to 42% in 2013, even if the higher education expansion should have pushed up the college share. This reflects that men's share of college educated workers in urban areas has achieved a saturation level and more young college graduates have to stay in rural areas.[8] The age structure is stable during the covered period shown by the stable averages and standard deviations.

Education-Occupation Match

The International Standard Classification of Occupations is commonly used to describe the occupation structure of a region. The version 2008, ISCO-08 is used in this paper. Figure 4.1 presents the major occupation groups, the correspondent skill levels, the occupation equivalents in CHIP, and the equivalent required education levels in CHIP.[9] In CHIP, the professional technicians are not as well classified as in ISCO-08

[6] See Card and Lemieux (2001) and Brunello (2010). Even if this issue may not be as severe as that in western countries considering that female labor force participation is relatively high in China(Meng, 2012), I focus on males for comparing results with existing literature mainly on western countries.

[7] I use nominal annual earnings in this paper, so the increase captures both real income growth and inflation. Using nominal earnings does not affect the results since the inflation index is canceled out in the estimates of college/non-college earnings gap.

[8] By comparing the share of college graduates in rural area between 2007 and 2013 using CHIP rural surveys, I indeed find this trend.

[9] The group, armed forces occupations, is not presented. It's difficult to categorize it into lower-

where they are further classified as the professionals, technicians/associated professionals. It brings in the difficulty to determine the exact required education level for professional technicians in CHIP because part of them may be required 4-year college degree or above while the others are required 3-year college degree. Considering that the managers' equivalent occupations in CHIP require mixed education levels as well, I categorize them as higher-skilled occupations for which the broadly defined college education is required. All other occupations are categorized as lower-skilled and the correspondent required education levels are secondary education or below.

Table 4.2 shows that the share of lower-skilled workers increased from 60 percent in 1995 to 75 percent in 2013. An explanation is that the demand for lower-skilled workers have increased by more than the demand for high-skilled workers in urban China as it became the world factory. Nevertheless, it's interesting that the increasing shares of lower-skilled occupations are coupled with the increasing shares of college-educated workers. It implies that there must have been more and more college educated workers who were placed into lower-skilled occupations.

Table 4.3 presents the trends in the share of higher-skilled occupations by education levels and age groups. Overall, 42 percent college educated workers and 13 percent non-college educated workers worked with higher-skilled occupations in 2013, 30 percent and 14 percent less compared to 1995. When examined by age groups, it's found that, for both education levels, the shares of higher-skilled occupations decreased among age group 25-34 by less than the other two older groups even though the supply of college educated workers increased most among age group 25-34. This fact implies that the education-occupation match may not be a main cause of the stagnated college premium for the young group, age 25-34. Table 4.4 presents the education-occupation match by another perspective. Among the higher-skilled occupations, more and more workers have a college degree, which reflects an increase in the education-occupation match. However, the mismatch increased meanwhile among the lower-skilled occupations since the share of college educated workers increased.

4.3 Decomposition Analysis

Methodology

Following the standard approach in the literature on cohort size effects, the samples are collapsed into cells based on single-year age (or multiple-year age groups) and survey year. Then the age (group) specific college earnings gap in each survey year is decomposed by the method proposed by Oaxaca (1973) and Blinder (1973). The Oaxaca-Blinder decomposition method is widely used for analyzing the earnings differential across genders, races, regions or any different groups of observations. The two groups in this paper are college educated workers and non-college educated workers. The overall earnings differential is decomposed into an explained component accounted by differences in characteristics and an unexplained component due to differences in coefficients. The decomposition is implemented by two steps. First, for survey year t, estimate college (c) and non-college(n) earnings regressions separately for individuals i among age group j (the subscripts i, j, t are suppressed to for simplicity):

$$y_c = \alpha_c + \beta_c H_c + \varepsilon_c \tag{4.1}$$

$$y_n = \alpha_n + \beta_n H_n + \varepsilon_n \tag{4.2}$$

where y_i is log annual earnings, α is a constant, H is a dummy variable that takes 1 for workers with higher-skilled occupations and 0 for workers with lower-skilled occupations, and β is the occupational premium.

Let a_c, a_n, b_c and b_n be the OLS estimates of α_c, α_n, β_c and β_n respectively, and denote the mean value with a bar over the variable. Then, since the OLS residuals have zero mean, we have:

$$\bar{y_c} - \bar{y_n} = (a_c - a_n) + (b_c - b_n)\bar{H_n} + b_c(\bar{H_c} - \bar{H_n}) \tag{4.3}$$

The first term at the right hand side is college earnings premium among the lower-skilled occupations since the constants represent the average log annual earnings of workers with lower-skilled occupations at different education level. The second term accounts for the differential due to the differences in occupational premiums across education groups. Put together, the first two terms measure the difference between the hypothesized earnings for a non-college educated worker in the college educated worker's earnings system and the true earnings. This part is usually called the unexplained differential in the literature.

The third term at the right hand side represents the earnings differential due to the differences in shares of higher-skilled occupations across education groups. The coefficient in this term b_c is the occupational earnings premium for college educated group which is usually called reference group. The selection of the reference group may affect the decomposition results. This can be shown by the following general form the decomposition,

$$\bar{y_c} - \bar{y_n} = (a_c - a_n) + (b_c - b^*)\bar{H_c} + (b^* - b_n)\bar{H_n} + b^*(\bar{H_c} - \bar{H_n}) \tag{4.4}$$

where the coefficient b^* is the OLS estimated occupational premium for the reference group.[10] The first term at the RHS is not affected, still measuring the college earnings premium among the lower-skilled occupations. The values of the other three terms depend on the value of b^*.

skilled or higher-skilled level, and the few armed forces workers are excluded in data.

[10]Table 4.1 summarizes several alternative decomposition methods following Oaxaca (1973) and Blinder (1973). They mainly concern the definition of b^*. And, in those studies, the groups studied are genders.

Decomposition Results

Since the primary goal of this paper is to examine how much the differences in occupational composition across education groups account for the magnitudes and trends of the overall college earnings premiums across age groups, the explained part, the last term at the RHS of equation (3.3), should be well examined with different reference groups. In this section, only the results based on b_c are reported because the results using other reference groups are similar to the ones reported here.

The decomposition is implemented to three age groups (age 25-34, 35-44, 45-54) in the five survey year 1995, 1999, 2002, 2007 (pooled together with 2008) and 2013. Table 4.5 not only presents the overall, explained and unexplained earnings gaps as decomposition analysis usually reports but also reports the estimated coefficients from the OLS regressions in the first step for examining the trends in them. Based on the results in table 4.5, the figures 4.2, 4.3, 4.4, and 4.5 provide more direct impressions about the earnings differentials and the decomposed components. Figure 4.2 depicts the divergent trends in the college earnings gaps across age groups between 1995 and 2013. The gaps for the younger group 25-34 stagnated since 1999 while the older group 45-54 kept increasing until 2013.

Figure 4.3 shows the explained gaps due to the differences in the shares of higher-skilled occupations. No similar divergent trend is revealed. They vary at low levels below 0.11. By equation (3.3), the explained earnings gap is the product of occupational shares differential and the occupational premium for college educated workers. To explain why the explained earnings gaps were low and stable, figures 4.6(a) and 4.6(b) present the shares of higher-skilled occupations across education groups and the differences in the shares. The shares of higher-skilled occupations among college educated workers were always higher than non-college educated workers, which demonstrates that the college education is helpful in obtaining better jobs. As one ages, the chance to get better jobs increases for both education groups. For all age groups, both college and non-college educated workers, the shares of higher-skilled occupations decreased between 1995 and 2013. It is surprising that the age group 45-54 decreased most considering that the relative supply of college educated workers among this old age group increased far less than the young group 25-34. This implies that supply effect on occupation reallocation is general across all age groups rather than contained within certain age group. Figure 4.6(b) depicts the differences in the share of higher-skilled occupations between college and non-college educated groups. Three age groups experienced similar decreasing trends. These decreasing share differentials, together with the increasing occupational premium b_c reported in table 4.5, resulted in the low and stable explained college earnings differential.

Figure 4.4 depicts the unexplained earnings gap due to the differences in occupational premium. No clear trends can be seen and all the gaps were negligible and even negative. This is mainly due to the small or even negative differences in the occupational premium between college and non-college groups, which can be seen from table 4.5.

Since the last component of the overall earnings gap is the unexplained part due to the constants, it must have contributed most to the overall earning gap. Figure

4.5 presents similar divergent trends as the those in overall college earnings gap in figure 4.2. We know that the constant measures the average earnings for lower-skilled occupations in each education group, therefore, it was the college earnings premium among lower-skilled occupations that accounted for most of overall earnings premium and significantly contributed to the divergent trends in the earnings premium across age groups.

The percentages of each component are presented in figure 4.7, 4.8, and 4.9. In summary, for all survey years and age groups, the differences in the share of higher-skilled occupations between college and non-college educated workers only explain a small part of college premium, 10%-30%. The part due to the higher-skilled occupational premium is negligible. Over 70% of the college premium is contributed by the college premium among the workers with lower-skilled occupations.

4.4 Cohort Size Effect on College Premium by Occupation Categories

It has been shown by figure 4.2 and 4.5, that the college earnings premium among the lower-skilled occupations also experienced divergent trends similar as the overall college premium. And, the negative cohort size effect on the overall college premiums has been demonstrated in chapter 3. Now, I turn to examine the negative cohort size effect among the lower-skilled occupations. But the standard model by Card and Lemieux (2001) does not model the formation of the college premium at occupation level. In this section, I develop the standard model by adding one more labor force aggregation step at occupation level to derive the college premium at occupation level.

Theoretical Model

I start with a Cobb-Douglas aggregate production function that has been widely used in the macro-growth literature:

$$Y_t = A_t L_t^{\alpha} K_t^{1-\alpha} \tag{4.5}$$

where subscript t indexes year, Y_t is aggregate output, A_t is total factor productivity, L_t is aggregate labor force input, K_t is physical capital input and α is the share of income allocated to labor force.

Following the existing literature on the trend of wage differentials by education (Katz and Murphy, 1992; Autor et al., 2008), I assume the labor force input L_t in equation 2.1 follows a CES aggregation of college and non-college laborïijŇ

$$L_t = [\sum_s (\theta_{st} L_{st}^{\rho})]^{1/\rho} \tag{4.6}$$

where subscript s indexes education level which takes c for college labor and n for non-college labor, θ_{st} is the technological efficiency parameter, and $-\infty < \rho \leq 1$ is a function of the elasticity of substitution σ_s between college and non-college labor force ($\rho = 1 - 1/\sigma_s$).

Next, I assume the labor force input L_{st} in equation (4.2) follows a CES aggregation of labor force with higher-skilled occupations and the labor force with lower-skilled occupations,

$$L_{st} = [\sum_{g}(\gamma_{sgt} L^{\tau}_{sgt})]^{1/\tau} \tag{4.7}$$

where subscript g indexes occupation levels which takes h for higher-skilled occupations and l for lower-skilled occupations, γ_{sgt} is the technological efficiency parameter, and $-\infty < \tau \leq 1$ is a function of the elasticity of substitution σ_g between higher-skilled and lower-skilled occupations ($\tau = 1 - 1/\sigma_g$).

To explain the divergent trends of the college premiums across age cohorts, following Card and Lemieux (2001), I relax the assumption of perfect substitution across age cohorts and further assume the labor force of each education level s and occupation category g is aggregated by age cohorts by CES functional formïijŇ

$$L_{sgt} = [\sum_{j}(\alpha_{sgj} L^{\eta}_{sgjt})]^{1/\eta} \tag{4.8}$$

where subscript j indexes age cohort, α_{sgj} is a time invariant relative efficiency parameter,[11], $-\infty < \eta \leq 1$ is a function of the elasticity of substitution σ_j across age cohorts ($\eta = 1 - 1/\sigma_j$), and L_{sgjt} is the size of labor force for each education-occupation-age-year group.

In this setup, assuming efficient utilization of labor force, I derive the profit-maximizing wage of an average worker with education level s, occupation category g, among age cohort j in year t as the value of corresponding marginal productivity in log form:

$$log(w_{sgjt}) = log(\Phi_t)+log(\theta_{st})+(\rho-\tau)log(L_{st})+log(\gamma_{sgt})+(\tau-\eta)log(L_{sgt})+log(\alpha_{sgj})+(\eta-1)log(L_{sgjt}) \tag{4.9}$$

where

$$\Phi_t = \alpha A_t K_t^{1-\alpha} L_t^{\alpha-\rho}$$

It is straightforward to derive the college premium by taking difference of the log wages between college and non-college labor force in terms of the following equation,

$$log(\frac{w_{cgjt}}{w_{ngjt}}) = log(\frac{\theta_{ct}}{\theta_{nt}})+(\rho-\tau)log(\frac{L_{ct}}{L_{nt}})+log(\frac{\gamma_{cgt}}{\gamma_{ngt}})+(\tau-\eta)log(\frac{L_{cgt}}{L_{ngt}})+log(\frac{\alpha_{cgj}}{\alpha_{ngj}})+(\eta-1)log(\frac{L_{cgjt}}{L_{ngjt}}) \tag{4.10}$$

where $log(\frac{\theta_{ct}}{\theta_{nt}})$ implies the year trend of the relative technological efficiency for college labor force, $log(\frac{L_{ct}}{L_{nt}})$ measures the relative size of aggregate college labor fore in year t, $log(\frac{\gamma_{cgt}}{\gamma_{ngt}})$ is the occupation specific trend in relative efficiency of college workers, $log(\frac{L_{cgt}}{L_{ngt}})$ measures the relative size of aggregate college labor fore with occupation g in year t, $log(\frac{\alpha_{cgj}}{\alpha_{ngj}})$ is the relative efficiency of college workers with occupation g in year t, and $log(\frac{L_{cgjt}}{L_{ngjt}})$ is the key variable of interest, the age specific relative size of college educated workers with occupation g in year t.

[11] I follow the assumption by Card and Lemieux (2001) that the relative efficiency parameter is

For a given occupation category, the first four terms at the right-hand-side of equation (4.6) capture the year trend of the college premium common for all age cohorts. The fifth term, $log(\frac{\alpha_{cgj}}{\alpha_{ngj}})$ captures the age cohort fixed effect. Thus, among occupation category g, the heterogeneous trends of the college premium across age cohorts should be due to the last term. And, the negative effect of age specific relative size on the college premium is expected unless workers are perfectly substitutable across age cohorts ($\eta = 1$).

Empirical Specification

Based on equation (4.6), for a given occupation category, the subscript g disappears, the regression specification can be written as

$$r_{jt} = F_t + F_j + \beta log(\frac{L_{cjt}}{L_{njt}}) + \varepsilon_{jt} \tag{4.11}$$

where r_{jt} is the college earnings premium for workers among given occupation category, age cohort j, and in year t, β measures the effect of cohort relative size of college educated workers on college premiums among the given occupation category. For all single-age and survey year cell, the college earnings premiums among higher-skilled and lower-skilled occupations can be obtained by implementing the decomposition cell by cell. Following Card and Lemieux (2001), I also record the standard errors of estimated cell specific college premiums. The corresponding inverse variances will be used as weights for the regression analysis to put more weight on those precisely estimated college premiums, and be used to construct goodness-of-fit tests for the null hypothesis that the relevant specification has no specification error.[12] The relative size of college educated workers is computed by taking log ratio of the number of college educated workers to non-college educated workers cell by cell. However, a simple OLS estimate of β may be biased. The specification is strictly based on the profit-maximizing wage functions which reflect only the demand side of the labor market, whereas the observed college premiums and the relative size of college educated workers represent the realized general equilibrium. Therefore, $log(\frac{L_{cjt}}{L_{njt}})$ may have been affected by the college premium through a supply channel. Thus, I use the predetermined variable, log ratio of the number of college degree holders to the number of non-college degree holders (including both employed and unemployed individuals, male and female) as an instrumental variable.

Results

In table 4.6, I present the estimated effects of age cohort specific relative size of college educated workers on the college premium for lower-skilled occupations. The

constant over time.

[12]Essentially, it tests whether the recorded variances of the estimated college premiums are significantly different from the variances of the residual in relevant specification. See Card and Lemieux

results by unweighted/weighted OLS estimation in columns 1 and 2 are not significantly different from zero. However as discussed in section 4.2, basic OLS estimate may suffer from the issue of simultaneous causation which makes it biased. I use the predetermined variable, log ratio of the number of college graduates to the number of non-college individuals (including both male and female, employed and unemployed), as an instrumental variable. The corresponding results are presented in column 3 and 4. The estimated relative size effects become significantly negative at 5 percent level, -0.188 by unweighted IV estimation and -0.153 by weighted IV estimation. The substantial increases in the magnitudes from OLS to IV estimates imply that the relative size of college workers might be positively affected by the college premium simultaneously. By weighted IV result, for lower-skilled occupations, holding year and age constant, a one unit increase in the relative size of college educated workers leads to about 15.3 percentage points decrease in the college premium. Table 4.7 reports the results for higher-skilled occupations. Again, the OLS estimates are not significant while the IV estimates are significantly negative at 5 or 10 percent level. The weighted IV estimate, -0.099 is not significantly different from the weighted IV result, -0.153, for the lower-skilled occupations.

4.5 Conclusion

This paper starts with a hypothesis that the divergent trends in the college earning premium across age groups may be partly driven by divergent trends in the education-occupation match across age groups in urban China between 1995 and 2013. Because China's higher education expansion has led to an increasing number of college-educated workers, especially among the younger cohorts. More and more workers among the younger cohorts may be placed into occupations which have typically been held by people with less education and lower earnings, which could result in the stagnated college earnings premium for the younger group. However, using Blinder-Oaxaca decomposition, it's found that, for all survey years and age groups, the differences in the share of higher-skilled occupations between college and non-college educated workers only explain a small part of college premium, 10%-30%. The part due to the occupational premium is negligible. Over 70% of the college premium is contributed by the college premium for the workers with lower-skilled occupations. These unexpected results reveal that the relative supply of college educated workers among certain age group has general effect rather than age group specific effect on the occupation reallocation.

Since the college premium among lower-skilled occupations account for over 70% of the overall college premium and is the driven force of the divergent trends by age groups, to explain this result, this paper further develops the model by Card and Lemieux (2001) to include one more labor force aggregation step at the occupation level. Using the specification based on the derived result, the cohort size effect on college premiums among the lower-skilled and higher-skilled occupations can be esti-

(2001) for details.

mated separately. Holding the age group and survey year constant, a one unit increase in log relative size of college educated workers is associated with about 15.3 percentage points decrease in college earnings premium among lower-skilled occupations, and 9.9 percentage points decrease among higher-skilled occupations.

Figures

Figure 4.1: Occupation Categories and Education levels

ISCO-08 major groups	Sill Level	Occupation in CHIP	Occupation Level	Education Level in CHIP	Education Level
1 Managers	3+4	Principals in State Agencies, Party Organizations, Enterprise and Public Institutions	Higher-Skilled	3-Year, 4-Year College and Above	Collge
2 Professionals	4	Professional Technicians		4-Year College and Above	
3 Technicians and Associated Professionals	3			3-Year College	
4 Clerical Support Workers	2	Clerk and Relevant Workers	Lower-Skilled	Junior, Senior and Vocational Senior Middle School	Non-College
5 Services and Sales Workers	2	Commercial and Service Workers		Junior, Senior and Vocational Senior Middle School	
6 Skilled Agricultural, Forestry and Fishery Workers	2	Agriculture, Forestry, Animal Husbandry, Fishery and Water Resource Producer		Junior, Senior and Vocational Senior Middle School	
7 Craft and Related Trades Workers	2	Manufacturing and Transportation Equipment Operators and Relevant Workers, Unskilled Workers		Junior, Senior and Vocational Senior Middle School	
8 Plant and Machine Operators, and Assemblers	2			Junior, Senior and Vocational Senior Middle School	
9 Elementary Occupations	1			Elementary School	

Figure 4.2: Trends in the College Premium across Age Groups

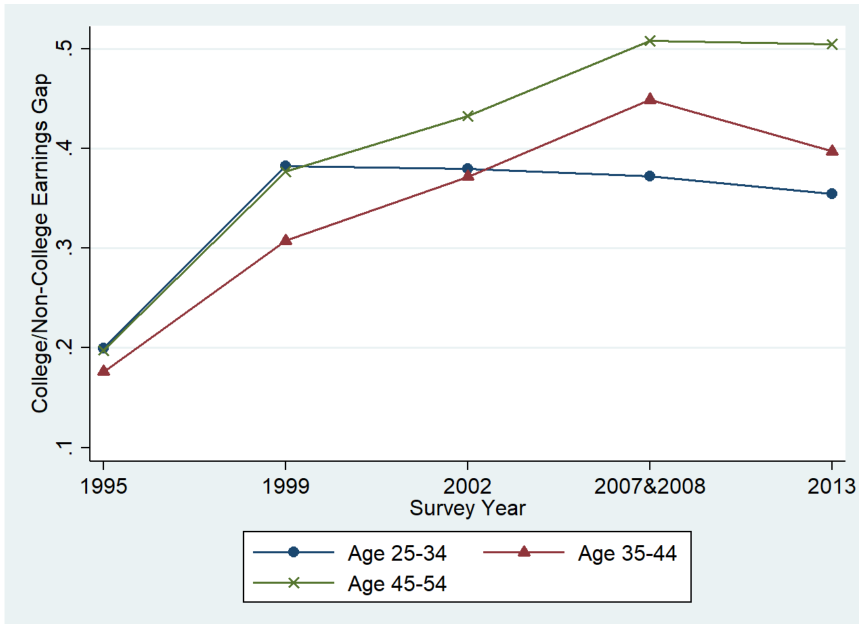


Figure 4.3: Trends in the Explained Earnings Gap across Age Groups

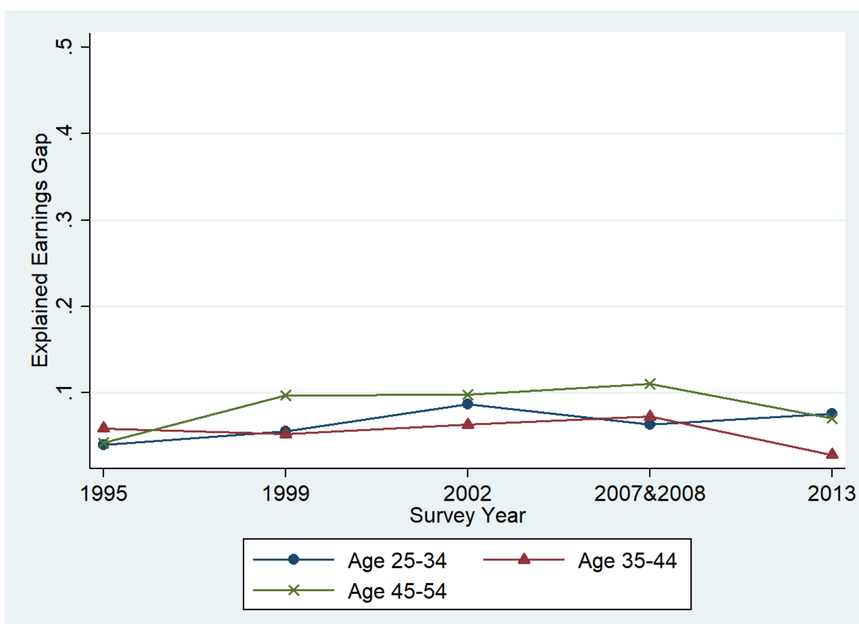


Figure 4.4: Trends in the Unexplained Earnings Gap Due to Occupational Premiums across Age Groups

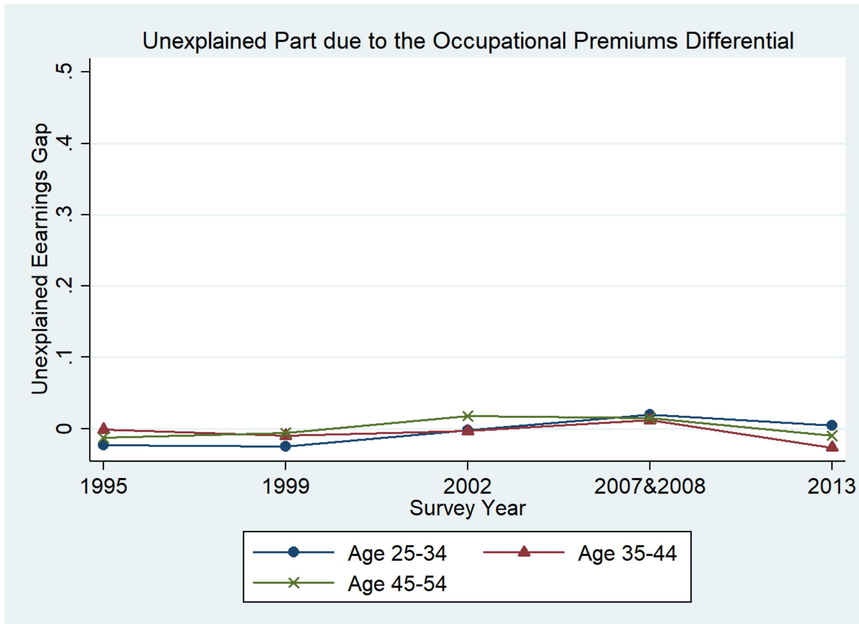


Figure 4.5: Trends in the Unexplained Earnings Gap Due to the Constants across Age Groups

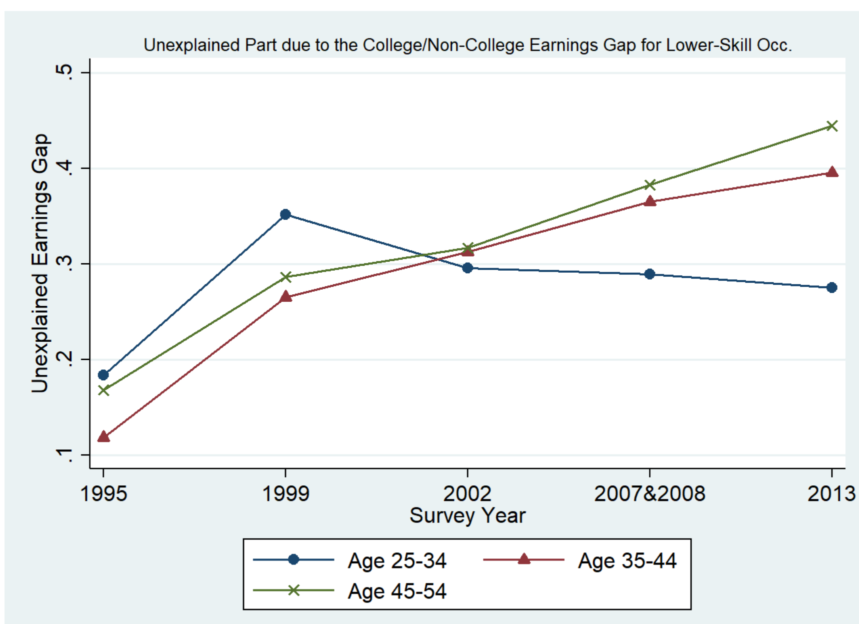
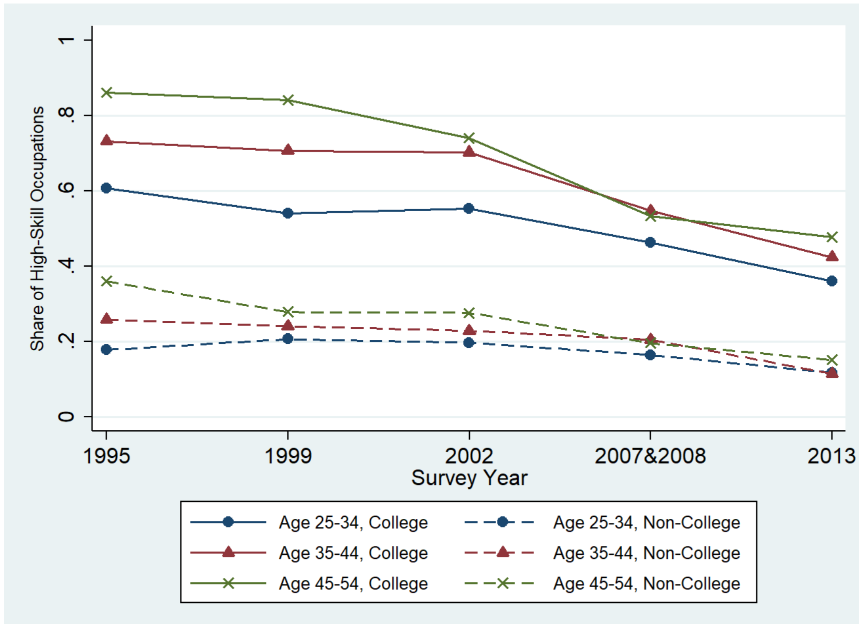


Figure 4.6: Trends in the Shares (Differential) of Higher-Skilled Occupations by Education and Age Groups

(a) Trends in the Shares of Higher-Skilled Occupations by Education and Age Groups



(b) Trends in the Shares Differential of Higher-Skilled Occupations by Age Groups

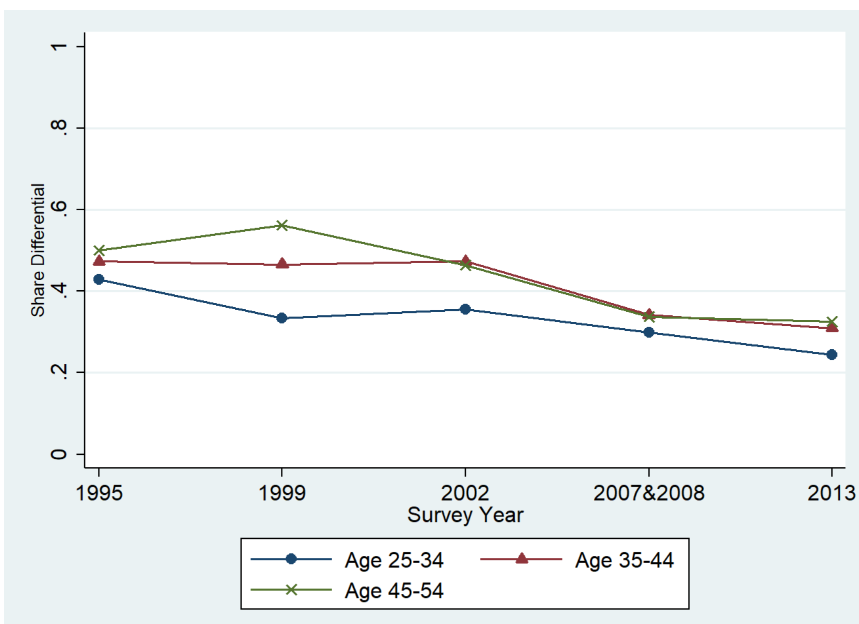


Figure 4.7: Trends in the Percentages of the Explained Earnings Gap By Age Groups

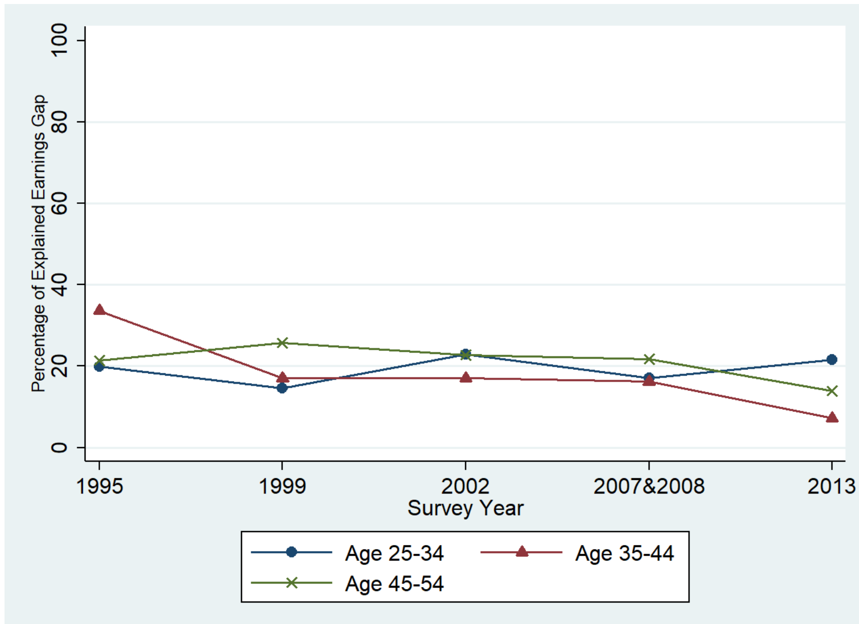


Figure 4.8: Trends in the Percentages of the Unexplained Earnings Gap Due to Occupational Premiums By Age Groups

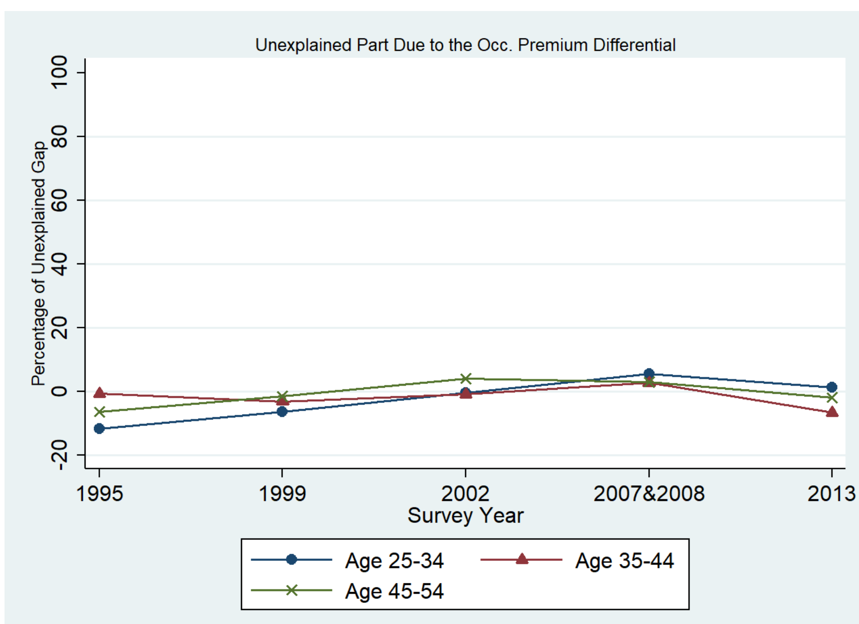
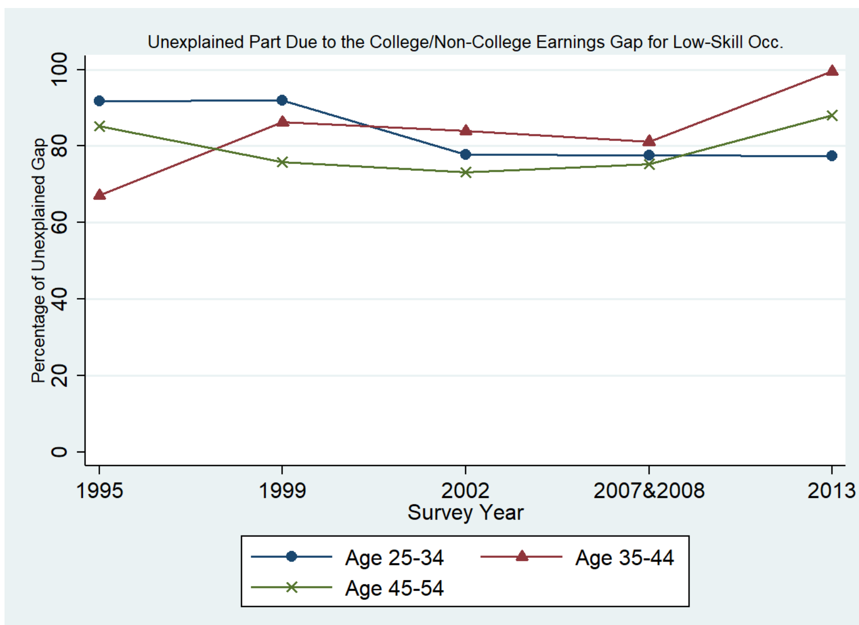


Figure 4.9: Trends in the Percentages of the Unexplained Earnings Gap Due to Constants By Age Groups



Tables

Table 4.1: List of Approaches to Determining b^*

Oaxaca (1973)	$b^* = b_m$	Male regression coefficients
Blinder (1973)	$b^* = b_f$	Female regression coefficients
Cotton et al. (1988)	$b^* = 0.5b_f + 0.5b_m$	Simple average of the coefficients in both groups
Reimers (1983)	$b^* = (\frac{N_{Male}}{N})b_m + (\frac{N_{Female}}{N})b_f$	Weighted average of the coefficients in both groups
Neumark (1988)	$b^* = b_p$	Coefficients from a pool regression

Table 4.2: Summary Statistics: Male Workers Only in CHIP Urban Data

CHIP	1995	1999	2002	2007	2013
Average Annual Earnings (Yuan)	6,374	8,690	11,159	27,174	36,316
College (%)	29	35	36	45	42
	(0.45)	(0.48)	(0.48)	(0.50)	(0.49)
Age	39.84	40.59	41.41	40.48	40.74
	(7.70)	(7.52)	(7.62)	(8.28)	(8.18)
Higher-Skilled Occ. (%)	40	41	40	34	25
Lower-Skilled Occ.(%)	60	59	60	68	75
Observations	4978	2754	4900	6461	4335

Table 4.3: Summary Statistics: Higher-Skilled Occ. Distribution by Education, Year, and Age Group

College-Educated Workers:	1995	1999	2002	2007	2013
All:					
Higher-Skilled Occ.(%)	72	70	67	51	42
N	1511	962	1747	2879	1803
Age 25-34:					
Higher-Skilled Occ.(%)	61	54	55	46	36
N	529	281	486	1125	623
Age 35-44:					
Higher-Skilled Occ.(%)	73	71	70	55	42
N	564	403	751	1118	655
Age 45-54:					
Higher-Skilled Occ.(%)	86	84	74	53	48
N	418	278	510	636	525
Non-College-Educated Workers:					
All:					
Higher-Skilled Occ.(%)	27	25	25	19	13
N	3682	1811	3165	3600	2538
Age 25-34:					
Higher-Skilled Occ.(%)	18	21	20	17	12
N	897	314	561	617	518
Age 35-44:					
Higher-Skilled Occ.(%)	26	24	23	21	12
N	1642	816	1116	1209	895
Age 45-54:					
Higher-Skilled Occ.(%)	36	28	28	20	15
N	1143	681	1488	1774	1125

Notes: Urban Samples from CHIP 1995, 1999, 2002, 2007,2008, and 2013.

Table 4.4: Summary Statistics: High-Skill Occ. Distribution by Education, Year, and Age Group

	1995	1999	2002	2007	2013
Higher-Skilled Occupations:					
All:					
College	0.52	0.60	0.60	0.68	0.69
N	2091	1123	1954	2172	1089
Age 25-34:					
College	0.67	0.70	0.71	0.84	0.79
N	481	217	380	624	286
Age 35-44:					
College	0.49	0.59	0.67	0.71	0.73
N	838	482	784	861	381
Age 45-54:					
College	0.47	0.55	0.48	0.49	0.59
N	772	424	790	687	422
Lower-Skilled Occupations:					
All:					
College	0.13	0.18	0.19	0.33	0.32
N	3102	1650	2958	4307	3252
Age 25-34:					
College	0.22	0.34	0.33	0.54	0.47
N	945	378	667	1118	855
Age 35-44:					
College	0.11	0.16	0.21	0.35	0.32
N	1368	737	1083	1466	1169
Age 45-54:					
College	0.07	0.08	0.11	0.17	0.22
N	789	535	1208	1723	1228

Notes: Urban Samples from CHIP 1995, 1999, 2002, 2007,2008, and 2013.

Table 4.5: Decomposition Results by Oaxaca-Blinder Decomposition

	1995	1999	2002	2007&2008	2013
Age 25-34:					
Overall	0.20***	0.38***	0.38***	0.37***	0.36***
Explained	0.04**	0.05**	0.09***	0.06***	0.08***
Unexplained(Occupational Premium)	-0.02*	-0.02	0	0.02	0
Unexplained(Constant)	0.18***	0.35***	0.29***	0.29***	0.28***
b_c	0.09**	0.17**	0.24***	0.21***	0.31***
b_n	0.23***	0.29***	0.26***	0.09	0.28***
a_c	8.62***	9.04***	9.26***	10.27***	10.46***
a_n	8.44***	8.69***	8.97***	9.98***	10.19***
Age 35-44:					
Overall	0.18***	0.31***	0.37***	0.45***	0.40***
Explained	0.06***	0.05**	0.06***	0.07***	0.03*
Unexplained(Occupational Premium)	0	-0.01	0	0.01	-0.03***
Unexplained(Constant)	0.12***	0.27***	0.31***	0.37***	0.40***
b_c	0.13***	0.11**	0.13***	0.21***	0.09*
b_n	0.13***	0.15***	0.15***	0.16***	0.32***
a_c	8.82***	9.20***	9.45***	10.39***	10.75***
a_n	8.70***	8.93***	9.14***	10.02***	10.35***
Age 45-54:					
Overall	0.20***	0.38***	0.43***	0.51***	0.51***
Explained	0.04	0.10**	0.10***	0.11***	0.07***
Unexplained(Occupational Premium)	-0.01	-0.01	0.01	0.02	-0.01
Unexplained(Constant)	0.17***	0.29***	0.32***	0.38***	0.45***
b_c	0.08	0.17**	0.21***	0.33***	0.22***
b_n	0.12***	0.19***	0.15***	0.25	0.28***
a_c	8.98***	9.27***	9.54***	10.31***	10.74***
a_n	8.81***	8.98***	9.22***	9.93***	10.29***

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Urban Samples from CHIP 1995, 1999, 2002, 2007,2008, and 2013.

Table 4.6: Basic Estimates for Effects of Age Specific Relative Size of College Workers on College Premiums: Lower-Skilled Occupations

Dependent Variable: College Premium	(1) OLS	(2) Weighted-OLS	(3) IV	(4) Weighted-IV
Log Relative Size	0.052	0.017	-0.188**	-0.153**
	(0.047)	(0.043)	(0.093)	(0.074)
Year Effects:				
1999	0.094**	0.150***	0.167***	0.227***
	(0.042)	(0.040)	(0.046)	(0.045)
2002	0.118***	0.171***	0.254***	0.271***
	(0.045)	(0.041)	(0.063)	(0.052)
2007	0.120*	0.180***	0.430***	0.399***
	(0.062)	(0.056)	(0.121)	(0.095)
2013	0.139**	0.208***	0.446***	0.425***
	(0.063)	(0.060)	(0.116)	(0.094)
F Statistic			21.29	34.36
χ^2(p-value)		112.60(0.55)		109.85(0.62)
Observations	150	150	150	150
R-squared	0.831	0.881	0.772	0.865

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable for all specifications is the college premiums by age and year for lower-skilled occupations. All specifications also include age fixed effects not reported. The instrumental variable for log relative size is log ratio of the number of college degree holders (including both male and female, employed and unemployed) to the number of non-college degree holders. Weights for specifications in columns 2 and 4 are the inverse sampling variance of estimated college premiums. Reference year is 1995.

Table 4.7: Basic Estimates for Effects of Age Specific Relative Size of College Workers on College Premiums: Higher-Skilled Occupations

Dependent Variable: College Premium	(1) OLS	(2) Weighted-OLS	(3) IV	(4) Weighted-IV
los	0.015	-0.053	-0.152**	-0.099*
	(0.094)	(0.044)	(0.066)	(0.051)
Year Effects:				
1999.year	0.118*	0.151***	0.168***	0.166***
	(0.059)	(0.043)	(0.047)	(0.039)
2002.year	0.188***	0.239***	0.239***	0.255***
	(0.055)	(0.032)	(0.040)	(0.029)
2007.year	0.287***	0.371***	0.399***	0.397***
	(0.079)	(0.040)	(0.049)	(0.038)
2013.year	0.185**	0.252***	0.312***	0.285***
	(0.080)	(0.059)	(0.072)	(0.058)
F Statistic			62.08	93.93
χ^2(p-value)		121.94(0.31)		121.06(0.33)
Observations	150	150	150	150
R-squared	0.678	0.821	0.653	0.819

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable for all specifications is the college premiums by age and year for higher-skilled occupations. All specifications also include age fixed effects not reported. The instrumental variable for log relative size is log ratio of the number of college degree holders (including both male and female, employed and unemployed) to the number of non-college degree holders. Weights for specifications in columns 2 and 4 are the inverse sampling variance of estimated college premiums. Reference year is 1995.

Bibliography

Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature 40*(1), 7–72.

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics 106*(4), 979–1014.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Ashenfelter, O. C. and A. B. Krueger (1994). Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review 84*(5), 1157–73.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics 90*(2), 300–323.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436–455.

Brunello, G. (2010). The Effects of Cohort Size on European Earnings. *Journal of Population Economics 23*(1), 273–290.

Byron, R. P. and E. Q. Manaloto (1990). Returns to education in china. *Economic Development and Cultural Change 38*(4), 783–796.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Card, D. (1999a). Can Falling Supply Explain the Rising Return to College for Younger Men âĂŸ? A Cohort-Based Analysis.

Card, D. (1999b). The causal effect of education on earnings. *Handbook of labor economics 3*, 1801–1863.

Card, D. and A. B. Krueger (1992). Does school quality matter? returns to education and the characteristics of public schools in the united states. *The Journal of Political Economy 100*(1), 1–40.

Card, D. and T. Lemieux (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

Carneiro, P. and J. J. Heckman (2002). The evidence on credit constraints in post-secondary schooling*. *The Economic Journal 112*(482), 705–734.

Carneiro, P. and S. Lee (2009). Estimating Distributions of Potential Outcomes Using Local Instrumental Variables With an Application to Changes in College Enrollment and Wage Inequality. *Journal of Econometrics 149*(2), 191–208.

Carneiro, P. and S. Lee (2011). Trends in Quality-Adjusted Skill Premia in the United States, 1960–2000. *The American Economic Review 101*(6), 2309–2349.

Chay, K. Y. and D. S. Lee (2000). Changes in Relative Wages in the 1980s Returns to Observed and Unobserved Skills and Black-White Wage Differentials. *Journal of Econometrics 99*(1), 1–38.

Cotton, F. A., G. Wilkinson, et al. (1988). *Advanced inorganic chemistry*, Volume 545. Wiley New York.

DuFLo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *The American Economic Review 91*(4), 795–813.

Fan, E., X. Meng, Z. Wei, and G. Zhao (2010). Rates of return to university education: The regression discontinuity design.

Feenstra, R. and G. Hanson (2001). Global Production Sharing and Rising Inequality: A Survey of Trade and Wages. Technical report, National Bureau of Economic Research.

Freeman, R. (1976). The overeducated american.

Freeman, R. B. and L. F. Katz (2007). *Differences and Changes in Wage Structures*. University of Chicago Press.

Gao, W. and R. Smyth (2015). Education expansion and returns to schooling in urban china, 2001–2010: evidence from three waves of the china urban labor survey. *Journal of the Asia Pacific Economy 20*(2), 178–201.

Ge, S. and D. T. Yang (2011). Labor market developments in china: A neoclassical view. *China Economic Review 22*(4), 611–625.

Giles, J., A. Park, and M. Wang (2015). The great proletarian cultural revolution, disruptions to education, and the returns to schooling in urban china.

Grogger, J. and E. Eide (1995). Changes in College Skills and the Rise in the College Wage Premium. *Journal of Human Resources*, 280–310.

Gustafsson, B., L. Shi, and H. Sato (2014). Data for studying earnings, the distribution of household income and poverty in china. *China Economic Review 30*, 419–431.

Gustafsson, B. A., L. Shi, and T. Sicular (2008). *Inequality and Public Policy in China*. Cambridge University Press.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*, 307–458.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of econometrics 30*(1-2), 239–267.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Johnson, E. N. and G. C. Chow (1997). Rates of return to schooling in china. *Pacific Economic Review 2*(2), 101–113.

Juhn, C., D. I. Kim, and F. Vella (2005). The Expansion of College Education in the United States: Is There Evidence of Declining Cohort Quality? *Economic Inquiry 43*(2), 303–315.

Kane, T. J. and C. E. Rouse (1995). Labor-market returns to two-and four-year college. *The American Economic Review 85*(3), 600–614.

Katz, L. F. et al. (1999). Changes in the Wage Structure and Earnings Inequality. *Handbook of Labor Economics 3*, 1463–1555.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kawaguchi, D. and Y. Mori (2016). Why Has Wage Inequality Evolved So Differently Between Japan and the US? The Role of the Supply of College-Educated Workers. *Economics of Education Review 52*, 29–50.

Kling, J. R. (2001). Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business & Economic Statistics 19*(3), 358–364.

Krueger, A. B. (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989. *The Quarterly Journal of Economics 108*(1), 33–60.

Lee, D. S. and D. Card (2008). Regression discontinuity inference with specification error. *Journal of Econometrics 142*(2), 655–674.

Lee, D. S. and T. Lemieuxa (2010). Regression discontinuity designs in economics. *Journal of economic literature 48*(2), 281–355.

Lemieux, T. and D. Card (2001). Education, earnings, and the âĂŸcanadian gi billâĂŹ. *Canadian Journal of Economics/Revue canadienne d'économique 34*(2), 313–344.

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

Li, S., J. Whalley, and C. Xing (2014). China's higher education expansion and unemployment of college graduates. *China Economic Review 30*, 567–582.

Lillard, L. A. (1977). Inequality: Earnings vs. Human Wealth. *The American Economic Review 67*(2), 42–53.

Liu, Z. (1998). Earnings, education, and economic reforms in urban china. *Economic development and cultural change 46*(4), 697–725.

Machin, S. and S. McNally (2007). Tertiary education systems and labour markets. *Education and Training Policy Division, OECD*.

Maurin, E. and S. McNally (2008). Vive la révolution! long-term educational returns of 1968 to the angry students. *Journal of Labor Economics 26*(1), 1–33.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Meng, X. (2012). Labor market outcomes and reforms in china. *The Journal of Economic Perspectives 26*(4), 75–101.

Miller, P., C. Mulvey, and N. Martin (1995). What do twins studies reveal about the economic returns to education? a comparison of australian and us findings. *The American Economic Review 85*(3), 586–599.

Mincer, J. A. (1974). Age and experience profiles of earnings. In *Schooling, experience, and earnings*, pp. 64–82. NBER.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources 23*(3), 279–295.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693–709.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *The American Economic Review 96*(1), 152–175.

Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World development 22*(9), 1325–1343.

Reimers, C. W. (1983). Labor market discrimination against hispanic and black men. *The review of economics and statistics*, 570–579.

Taber, C. R. (2001). The rising college premium in the eighties: Return to college or return to unobserved ability? *The Review of Economic Studies 68*(3), 665–691.

Taubman, P. and T. Wales (1974). *Higher education and earnings.* McGraw-Hill Book.

Wang, L. (2012). Economic Transition and College Premium in Urban China. *China Economic Review 23*(2), 238–252.

Wang, X., B. M. Fleisher, H. Li, and S. Li (2014). Access to college and heterogeneous returns to education in china. *Economics of Education Review 42*, 78–92.

Wang, X. and J. Liu (2011). ChinaâĂŹs higher education expansion and the task of economic revitalization. *Higher Education 62*(2), 213–229.

Welch, F. (1979). Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust. *Journal of Political Economy 87*(5, Part 2), S65–S97.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Zhang, J., Y. Zhao, A. Park, and X. Song (2005). Economic returns to schooling in urban china, 1988 to 2001. *Journal of comparative economics 33*(4), 730–752.

Vita

Chenxu Hu

EDUCATION	**University of Kentucky**, Lexington, KY, US
	M.Sc. Economics, December 2014
	Shanghai University of Finance and Economics, China
	M.A. Economics, July 2013
	Beihang University, China
	B.E. Mechanical Engineering, July 2007
EMPLOYMENT	**University of Kentucky**
	Instructor (2017)
	ECO391 Business and Economics Statistics (3 sections)
	Teaching Assistant (2014-2015)
	-ECO 471 International Trade
	-ECO 491 Applied Econometrics
	-ECO 410 Current Issues in Economics
	Shanghai University of Finance and Economics
	Research Assistant (2012)
	Foxconn Technology Group
	Equipment Engineer (2007-2009)
CONFERENCE	**SEA 87th Annual Meetings**, November 2017
	Chinese Economists Society North American Conference, March 2017
	International Symposium on Human Capital and Labor Markets, December 2016
	Kentucky Economic Meeting, October 2016
HONORS AND AWARDS	China Scholarship Council (2013-2017)
	Gatton Fellowship (2014-2016)